Applications and Industry®

July 1960

Transactions Papers

General Applications Division

60-196 Voltage Regulation in Aircraft Electric Systems.....Krausz, Kahle . . . 121
60-195 Transistorized D-C Voltage Regulator.....Corey, Hansen . . . 128

Power Division

60-127 Brushless D-C Excited Rotating Field Synchronous Motor...Rosenberry . . . 136
60-166 Pull-In Criterion for Reluctance Motors.....Douglas . . . 139

Industry Division

60-62 Vibratory and Rotating-Wheel Gyroscopic Rate Indicators.....Newton . . . 143
60-16 Characteristics of Centrifugal Pumps and Compressors..Wiegand, Eddy . . . 150
60-52 Static Switching of Oil-Fired Combustion Turbines...Carmack, Smith . . . 157
60-116 Nonlinear Compensator for Feedback System....Athanassiades, Smith . . . 167
60–117 Stability Analysis of Dual-Mode Servomechanisms.....Gibson, McVey . . . 173
60-34 Forward Voltage Drop and Power Loss in Silicon Rectifiers.....Luft . . . 179
60-111 Systematic Approximation to Partition Method.....Ku, Wolf, Dietz . . . 183
60-241 Measurement of Temperature in Welding Arcs.....Ludwig . . . 191
60-240 Analysis of Transfer in Gas-Shielded Welding Arcs.....Greene . . . 194
60-140 New Welder Busway Distribution System.....Fisher, Dailey . . . 203
60-118 Servo with Backlash, Coulomb Friction, and Stiction....Pastel, Thaler . . . 215
60-115 Input-Output Cross-Correlation Functions.....Leland . . . 219
59-1296 Optimization of Adaptive Function by Z-Transform Method.....Chang . . . 223
Errata..... . . . 231
Conference Papers Open for Discussion.....See 3rd Cover

NUMBER 49
Published Bimonthly by
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Instrumentation Division

60-158 The Combi-System—New Concepts in Digital Data Processing...Schwab . . . 193
60-114 Precision Voltage Reference for Industrial Recorders.....Robinson . . . 197
60-135 Thermal Voltage Converters.....Hermach, Williams . . . 200

Industry Division

60-30 D-C Magnetic Crane Hoist Control.....Myles, Davies, Srnka . . . 207

Communication Division

60-44 An All-Transistorized Trunk Carrier System.....Gardner . . . 212
60-126 A Wired-Memory Translator with Shared Access.....Pace, Ostendorf . . . 216
60-142 Radio Communication with Orthogonal Time Functions.....Harmuth . . . 221
60-27 The Effect of Delay Distortion on Data Transmission.....Mertz . . . 228
60-233 A 240-Line Fully Electronic Telephone Switchboard.....Goudet . . . 232
60-46 Test Set for Measurement of Envelope Delay Distortion.....Codd . . . 241
60-161 Intermodulation Effects in FM and PM Systems.....Love . . . 245
59-1076 Transmission of Information by Orthogonal Time Functions..Harmuth . . . 248

Science and Electronics Division

60-129 Magnetic Analog with Spiral Search Coils.....Moore . . . 256
60-74 Electron Tube or Semiconductor Device?.....Scheneman, Waldorf . . . 264
60-143 Operation of the Magnetic Second-Harmonic Modulator.....Jalbert . . . 268
Inductors Biased with Permanent Magnets.....Ludwig
60-198 Part I. Theory and Analysis..... . . . 273
60-199 Part II. Design and Synthesis..... . . . 278
60-247 Dynamic Optimizing Control of the Continuous Process....Sandelien . . . 291
60-197 A Method for the Design of Holding Electromagnets.....Ludwig . . . 300
60-15 Measurement and Influence of Surface Charges.....Reynolds . . . 310
59-120 Transformer Design by Digital Computers.....Deise, Etchison, Lee . . . 314
60-202 Synthesis of Linear Filters.....Gilbert, Otterman . . . 323
60-150 Eddy-Current Brake for Sodium-Cooled Nuclear Power Reactor..Baker . . . 330
60-203 Difference Equations and Their Applications.....Pipes . . . 333
59-1190 Magnetic-Tape-to-Paper-Tape Converter.....Ringer, Mintzer . . . 339
60-201 Polarized Electromagnet and Vibrator Drive Systems.....Nunn . . . 346

(See inside back cover)

Note to Librarians. The six bimonthly issues of "Applications and Industry," March 1960–January 1961, will also be available in a single volume (no. 79) entitled "AIEE Transactions—Part II. Applications and Industry," which includes all technical papers on that subject presented during 1960. Bibliographic references to Applications and Industry and to Part II of the Transactions are therefore equivalent.

Applications and Industry. Published bimonthly by the American Institute of Electrical Engineers, from 20th and Northampton Streets, Easton, Pa. AIEE Headquarters: 33 West 39th Street, New York 18, N. Y. Address changes must be received at AIEE Headquarters by the first of the month to be effective with the succeeding issue. Copies undelivered because of incorrect address cannot be replaced without charge. Editorial and Advertising offices: 33 West 39th Street, New York 18, N. Y. Nonmember subscription $8.00 per year (plus 75 cents extra for foreign postage payable in advance in New York exchange). Member subscriptions: one subscription at $5.00 per year to any one of three divisional publications: Communication and Electronics, Applications and Industry, or Power Apparatus and Systems: additional annual subscriptions $8.00 each. Single copies when available $1.50 each. Second-class mail privileges authorized at Easton, Pa. This publication is authorized to be mailed at the special rates of postage prescribed by Section 132.122.

Printed in United States of America

Number of copies of this issue 5,100

oltage Regulation in Modern Aircraft Electric Power Systems

ALFRED KRAUSZ
MEMBER AIEE

H. A. KAHLE
ASSOCIATE MEMBER AIEE

HE PURPOSE of this paper is to discuss the general topic of voltage ılation for modern a-c aircraft and sile electric systems. Both steady-e and transient conditions will be con-red. Since there are many types of ipment which utilize a-c electric ver and since their number and com-xity are growing all the time, such a cussion should prove of merit to both user and the supplier of the electric ver.

The first portion of this discussion deals h the need for accurate voltage regula-ı and low transient-voltage deviations. sent design requirements for various itary and commercial installations are iewed. The second part gives an lysis of the synchronous generator n the point of view of the feedback trol engineer who has to design the tage regulator. The extreme com-xity of even the elementary syn-onous machine when viewed as part of ervosystem will be demonstrated.

Finally, various types of rotating and tic exciters as they influence the tran-ıt response of the system will be con-ered. In keeping with the purpose of s paper, fundamental considerations l be stressed without giving numerical ign details.

ed for Voltage Regulation

For purposes of this discussion of volt-e regulation, the various uses of a-c wer aboard aircraft, missiles, or space nicles may be grouped as follows:

To provide power for driving motors, mps, etc.

To supply a-c operated instruments such gyros, synchros, induction potentiome-s, etc.

To provide d-c power by means of trans-mer rectifiers to transistor and vacuum-be amplifiers and various control circuits.

per 60-196, recommended by the AIEE Aero-ce Transportation Committee and approved by AIEE Technical Operations Department for sentation at the AIEE Winter General Meeting, w York, N. Y., January 31–February 5, 1960. nuscript submitted April 20, 1959; made ilable for printing December 11, 1959.

FRED KRAUSZ, formerly with Jack & Heintz, ., Cleveland, Ohio, is with Space Technology oratories, Inc., Los Angeles, Calif.; H. A. HLE is with Jack & Heintz, Inc., Cleveland, Ohio.

4. To furnish light and heat.

5. To provide primary a-c power to computers and radar sets.

The necessary characteristics of the primary a-c power differ for each of these applications. The basic question arises as to whether it is better to generate the primary a-c power with sufficiently good characteristics to satisfy the most critical load, or whether to take whatever a-c power can be obtained easily and then refine it before application to the critical load. This question must be answered for each new system after a thorough analysis of the loads, the mission profile, and the over-all economics of the problem. It is a problem that absorbs a great many engineering hours in each new application and cannot be answered without extensive trade-off studies. For military aircraft applications the usual specification for the generator and regulator is *MIL-G-6099A*. This allows steady-state voltage limits of 111.5–118.5 at the point of regulation with balanced loads. Limits on transient response are given in terms of the voltage envelope which allows a maximum overshoot of 36% upon removal of load. Other pertinent requirements involve voltage modulation of 1% crest factor of 1.414±10%, and a 3% limit on any harmonic. New military utilization equipment is generally designed to tolerate *MIL-G-6099A* limits but there are certain critical applications where better transient response and better voltage regulation would be an advantage. The following deserve consideration:

1. In many instances semiconductor components such as rectifiers, diodes, controlled rectifiers, and switching transistors are used in a-c circuits. The peak inverse voltage (PIV) of these components must be sufficiently high at the most critical operating temperature to withstand the peak 1-cycle transient voltage. Lower transient overshoots on the primary a-c bus would allow use of semiconductors with lower PIV, and hence lower cost. Since the power-handling capability of the semiconductor in many of these applications is directly proportional to the PIV, a better transient response would also give maximum utilization of the power-handling capacity of these components.

2. For most amplifiers, regulated d-c supply voltage is required. Since the forward resistance of semiconductor rectifiers can be made very small, it is possible to design transformer units with sufficiently good regulation for many applications provided the a-c input voltage is kept constant. Close steady-state voltage regulation of the primary a-c supply, therefore, eliminates the need for regulated transformers in most applications.

3. The thermal characteristics of many a-c devices are designed to very close limits. In some applications, for instance when heating of gyromotors is critical, close control of the rms value of the motor supply voltage is essential.

4. In certain computers and radar equipment the a-c system is used for timing and waveform selection processes. Excessive transients in the a-c supply can cause computational errors that cannot be tolerated.

5. Amplitude modulation of the a-c supply can usually not be tolerated because the amplitude modulation will appear as an a-c signal superimposed on the output of a transformer unit which it supplies. This is especially serious if the modulation is of a frequency sufficiently low to affect the performance of an autopilot which constitutes a load.

The above considerations all lead to the conclusion that both transient and steady-state voltage deviations should be as small as possible and that voltage modulation must be avoided. Additional arguments can be advanced in favor of low waveform distortion.

The next questions then are: Just how good can the regulation and transient response be made in a practical system and what are the general guide lines that must be followed in achieving optimum performance? To answer these questions, one must first consider the transient characteristics of the generator which supplies the a-c power and then extend the analysis to the inclusion of the generator excitation system which encompasses the voltage regulator. Questions as to the effect of the constant-speed drive on the characteristic of the bus voltages are outside the scope of this paper and will be deferred for discussion at a later date.

Nomenclature

i = current
H = transfer function of damper
R,r = per-unit resistance
T = time constant
T_{do} = open-circuit field time constant
p = per-unit time differential operator $d/d(\omega_0 t)$
$p\theta$ = per-unit speed (equals one at rated speed ω_0)
x = per-unit reactance
Z = per-unit load impedance
γ = saturation factor
ψ = flux linkage
ω = angular frequency
λ = leakage factor

SUBSCRIPTS

a = armature
al = armature leakage
d = direct axis
ff = self-inductance
fl = field leakage
f = field
kd = direct-axis damper
kk = damper self-inductance
kl = damper leakage inductance
kq = quadrature-axis damper
L = load
md = direct-axis magnetizing
mq = quadrature-axis magnetizing
q = quadrature axis

Synchronous Machine Equations

Park[1] has formulated the equations of a synchronous machine in a stationary reference frame, using the 2-axis components o, d, q. In this form, the resulting quasi-holonomic equations stand for a fundamental machine from which the equations for nearly all other machines can be derived. The currents and voltages of a synchronous machine are resolved along these stationary reference axes. The main field is placed in the direct axis. Any damper winding is resolved into two component coils along the d- and q-axis.[2] The machine equations expressed in this reference frame are as follows:

$$
\begin{aligned}
v_f &= r_f i_f + p\psi_f \\
v_d &= -r_a i_d + p\psi_d - \psi_q p\theta \\
v_q &= -r_a i_q + p\psi_q + \psi_d p\theta \\
v_{kd} &= r_{kd} i_{kd} + p\psi_{kd} \\
v_{kq} &= r_{kq} i_{kq} + p\psi_{kq}
\end{aligned} \tag{1}
$$

The flux linkages of the field, armature, and damper windings in the d- and q-axis are given by

$$
\begin{aligned}
\psi_f &= \frac{N_f}{N_d}\psi_{md} + L_{fl} i_f \\
\psi_d &= \psi_{md} - L_{al} i_d \\
\psi_q &= \psi_{mq} - L_{al} i_q \\
\psi_{kd} &= \frac{N_{kd}}{N_d}\psi_{md} + L_{kd} i_{kd} \\
\psi_{kq} &= \frac{N_{kq}}{N_q}\psi_{mq} + L_{kq} i_{kq}
\end{aligned} \tag{2}
$$

The terms ψ_{md} and ψ_{mq} refer to mutual flux linkages in the direct and quadrature axis. Reducing all quantities to per-unit quantities referred to the armature allows one to express the flux linkages in terms of per-unit current and inductances as:

$$
\begin{aligned}
\bar{\psi}_{md} &= \bar{L}_{md}(\bar{\imath}_f - \bar{\imath}_d + \bar{\imath}_{kd}) \\
\bar{\psi}_{mq} &= \bar{L}_{mq}(-\bar{\imath}_q + \bar{\imath}_{kq})
\end{aligned} \tag{3}
$$

Since from this point on all quantities are per unit, the bar which denotes "per unit" will be omitted. Also, the differential operator p is redefined as $p = d/d(\omega_0 t)$. The inductance of any winding is equal to the sum of a magnetizing inductance and a leakage inductance, e.g., the self-inductance of the direct-axis damper winding is equal to $x_{kkd} = x_{md} + x_{kd}$.

Substitution of equations 2 and 3 into equation 1 yields the machine equations in terms of per-unit currents and voltages. These equations, in form of a matrix, and taken at base speed, are as follows:

$$
\begin{Vmatrix} v_f \\ v_d \\ v_q \\ v_{kd} \\ v_{kq} \end{Vmatrix} = \begin{vmatrix} r_f + px_{ff} & -px_{md} & 0 & px_{md} & 0 \\ px_{md} & -r_a - px_d & X_q & px_{md} & -x_{mq} \\ x_{md} & -x_d & -r_a - px_q & x_{md} & px_{mq} \\ px_{md} & -px_{md} & 0 & r_{kd} + p(x_{md} + x_{kd}) & 0 \\ 0 & 0 & -px_{mq} & 0 & r_{kq} + p(x_{mq} + x_{kq}) \end{vmatrix} \times \begin{Vmatrix} i_f \\ i_d \\ i_q \\ i_{kd} \\ i_{kq} \end{Vmatrix} \tag{4}
$$

The load on the generator can be represented by the following matrix:

$$
\begin{Vmatrix} v_d \\ v_q \end{Vmatrix} = \begin{vmatrix} r_L & -x_L \\ x_L & r_L \end{vmatrix} \times \begin{Vmatrix} i_d \\ i_q \end{Vmatrix} \tag{5A}
$$

or in inverse form:

$$
\begin{Vmatrix} i_d \\ i_q \end{Vmatrix} = \begin{vmatrix} r_L/Z^2 & x_L/Z^2 \\ -x_L/Z^2 & r_L/Z^2 \end{vmatrix} \times \begin{Vmatrix} v_d \\ v_q \end{Vmatrix} \tag{5B}
$$

with $Z^2 = |r_L^2 + x_L^2|$.

Signal Flow Diagram

The two matrices, equations 4 and 5, describe the behavior of a synchronous generator with any type of series resistive-inductive load. The problem is to solve for v_d and v_q for any suddenly applied load change and to determine how v_f should change to keep v_d and v_q between prescribed limits during the ensuing transient. The equations are far too complicated to do so and still allow an insight into machine characteristics. Block diagrams have been developed[2] for computer simulation. Though these block diagrams are convenient for simulation work they lack visibility of the interaction of the different variables. A signal flow diagram[3,4] provides a better picture and more details than a block diagram and also retains visibility of the flow of signals through the machine. The signal flow diagram of Fig. 1 represents the interdependency of the different currents and voltages in a synchronous generator supplying power to a resistive-inductive series load according to equations 4 and 5. The value of a variable at each nodal point is given by the different signals entering the node. Each signal is multiplied by the transmittance along its signal path.

The damper windings have been represented by a transfer function H and by the damper time constant T_{kd} which is defined as follows:

$$
v_{kd} = (i_f - i_d)px_{md} + i_{kd}[r_{kd} + p(x_{md} + x_{kd})] =
$$

since no voltage source is present. From this the damper current is

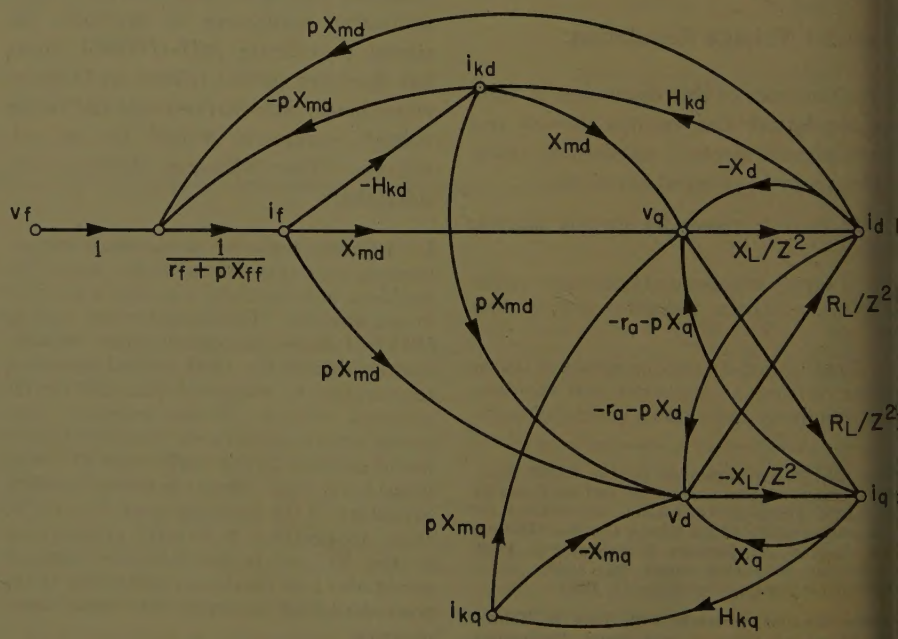


Fig. 1. Signal flow diagram of synchronous generator

$$= -(i_f - i_d)\frac{px_{md}}{r_{kd}+p(x_{md}+x_{kd})}$$

$$= \frac{px_{md}/r_{kd}}{1+pT_{kd}}(i_d - i_f) = (i_d - i_f)H_{kd} \quad (6)$$

h $T_{kd} = (x_{md}+x_{kd})/r_{kd} = x_{kkd}/r_{kd}$, the
-unit time constant of the damper
ding.

imilar equations hold for the q-
per winding:

$$= i_q \frac{px_{mq}/r_{kq}}{1+pT_{kq}} = i_q H_{kq} \quad (7)$$

t as it is possible to manipulate the
tem of simultaneous equations 4 and
o certain rules[3] allow the reduction of
signal flow diagram to a less com-
ated form and nodal points can be
ted or removed completely. By use
this technique, the complex flow dia-
m of Fig. 1 reduces step by step to the
complicated one of Fig. 2. The nodal
nts for the two damper windings
e been eliminated, but their effect is
luded in the transmittances along the
al paths. The field voltage v_f drives
eld current i_f which in turn generates
drature- and direct-axis voltages v_q
d v_d. The load currents are produced
these voltages and their effect on the
chine is also indicated.

duction of Signal Flow Diagram

he signal flow diagram of the genera-
as shown in Fig. 2 can be reduced to a
re practical form by manipulation of
transmittances along the different
al paths. The transfer functions of
two damper windings equal H_{kd} and
and represent the ratio of current in
damper to the sum of the remaining
rents in the same axis (equations 6
1 7). The term $(1-H_{kd})$ which ap-
rs throughout Fig. 2 is equal to:

$$H_{kd} = \frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}} \quad (8)$$

ere

$=x_{kd}/(x_{md}+x_{kd})$

l

$=(x_{md}+x_{kd})/r_{kd}\,\omega_0$

$_{kd}$ represents the percentage leakage
the damper winding. Similarly, one
s for the quadrature-axis damper:

$$H_{kq} = \frac{1+\lambda_{kq}T_{kq}}{1+pT_{kq}} \quad (9)$$

e transmittance for the field circuit
be written in terms of field and damper
e constants and leakage factors. A
d leakage factor can be defined as for
damper, given by:

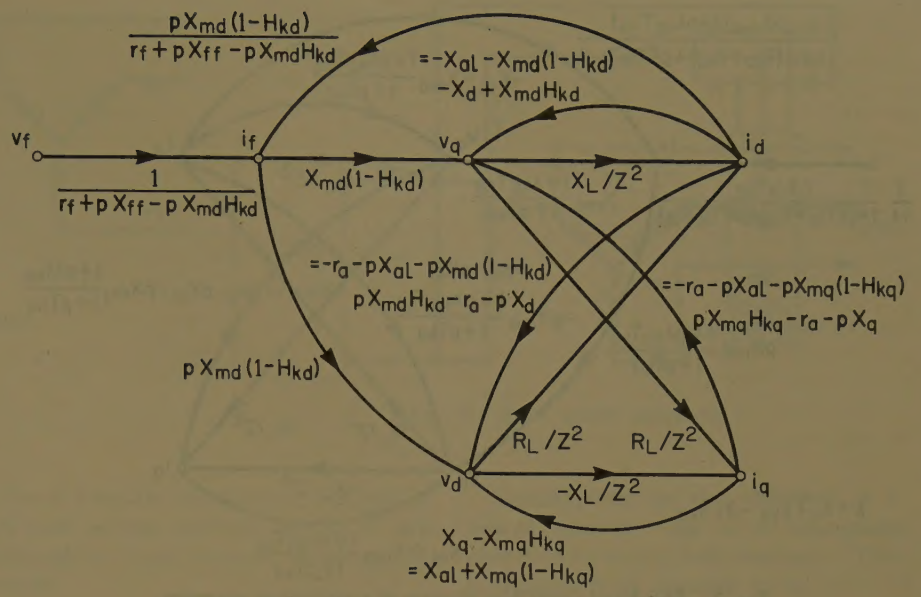


Fig. 2. Signal flow diagram, damper currents eliminated

$$\lambda_f = \frac{x_{fl}}{x_{md}+x_{fl}}$$

From Fig. 2 the field transmittance is equal to:

$$\frac{1}{r_f+px_{ff}-px_{md}H_{kd}} = \frac{1}{r_f+px_{ff}-p^2x^2_{md}/(1+pT_{kd})}$$

$$= \frac{1}{r_f}\,\frac{1+pT_{kd}}{1+p(T_{do}+T_{kd})+p^2[x_{md}x_{fl}+x_{md}x_{kd}+x_{fl}x_{kd})/r_f r_{kd}}$$

$$= \frac{1}{r_f}\,\frac{1+pT_{kd}}{1+p(T_{do}+T_{kd})+p^2T_{do}T_{kd}(\lambda_f+\lambda_{kd}-\lambda_f\lambda_{kd})}$$

The quadratic term in the denominator represents time constants associated with the leakage reactances of field- and direct-axis damper winding, which are usually small.

Using the above equations, the signal flow diagram reduces to that of Fig. 3, which represents the synchronous generator in a quite visible form. A computer block diagram can be derived from the signal flow diagram. However, this will not be discussed in the present paper. The functional relationship between the different variables can also be given in the form of a set of equations which now can be obtained directly from the signal flow diagram. These equations for the synchronous generators are

$$v_d = -\left[r_a+px_{al}+px_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}\right]i_d+\left[x_{al}+x_{mq}\frac{1+p\lambda_{kq}T_{kq}}{1+pT_{kq}}\right]i_q+px_{mq}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}\,i_f$$

$$v_q = -\left[x_{al}+x_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}\right]i_d-\left[r_a+px_{al}+px_{mq}\frac{1+p\lambda_{kq}T_{kq}}{1+pT_{kq}}\right]i_q+X_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}\,i_f \quad (10)$$

$$v_f = -px_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}\,i_d-\frac{1+p(T_{do}+T_{kd})+p^2T_{do}T_{kd}[\lambda_{kd}+\lambda_f-\lambda_{kd}\lambda_f]}{1+pT_{kd}}\,r_f i_f$$

This set of equation describes machine currents and voltages in terms of machine parameters. In terms of time constants for the various equivalent d- and q-axis circuits, these equations can be written as:

$$v_d = -\frac{x_d+x'_{kd}T_{kd}p}{1+pT_{kd}}\,pi_d+\frac{x_q+x'_{kq}T_{kq}p}{1+pT_{kq}}\,i_q+\frac{(1+pT_{kLd})x_{md}p}{1+pT_{kd}}\,i_f \quad (11)$$

$$v_q = -\frac{x_d+x'_{kd}T_{kd}p}{1+pT_{kd}}\,i_d-\frac{x_q+x'_{kq}T_{kq}p}{1+pT_{kq}}\times pi_q+\frac{(1+pT_{kLd})x_{md}}{1+pT_{kd}}\,i_f$$

$$v_f = -\frac{(1+pT_{kLd})x_{md}}{1+pT_{kd}}\,pi_d+\frac{1+p(T_{do}+T_{kd})+p^2(T_{do}T_{kd}-T'_{fd}T'_{kd})}{1+pT_{kd}}\,i_f r_f$$

where $T_{kLd} = (x_{kkd}-x_{md})/r_{kd} = \lambda_{kd}T_{kd}$ is the time constant associated with the

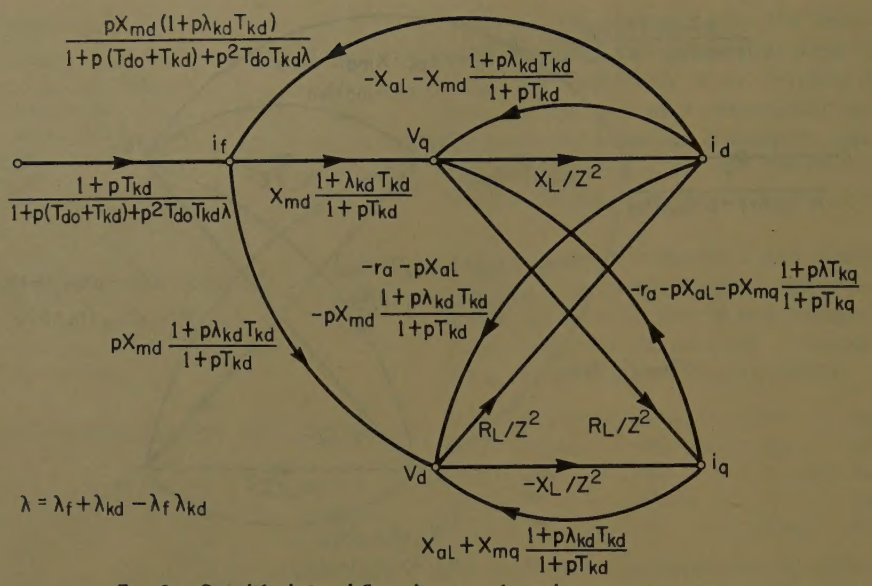


Fig. 3. Simplified signal flow diagram of synchronous generator

damper leakage reactance. T'_{fd} and T'_{kd} are defined as time constants of the magnetizing reactance equal to

$$T'_{fd}=x_{md}/r_f$$

and

$$T'_{kd}=x_{md}/r_{kd}$$

In addition, transient reactances defined as follows have been used in the foregoing equations:

$$x'_d=x_d-x^2_{md}/x_{ff}$$

$$x'_{kd}=x_d-x^2_{md}/x_{kkd}$$

$$x'_{kq}=x_q-x^2_{mq}/x_{kkq}$$

Discussion of Machine Characteristics

The case of a generator without load shall be considered first since it reduces the complexity of equations 10 or 11 and of the signal flow diagram of Fig. 3 considerably. Since the load reactance and resistance are infinite, all signals leaving the nodes v_q and v_d are zero. As a result, there will be no current in direct or quadrature axis. The equations for no load are

$$v_d=px_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}i_f$$

$$v_q=x_{md}\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}i_f \qquad (12)$$

$$v_f=\frac{1+p(T_{do}+T_{kd})+p^2T_{do}T_{kd}\times(\lambda_f+\lambda_{kd}-\lambda_f\lambda_{kd})}{1+pT_{kd}}r_fi_f$$

The direct-axis damper time constant governs the response of the machine with respect to changes in field current. The field current produces the quadrature voltage v_q according to the slope x_{md} of the no-load saturation curve. However, response of v_q with respect to changes in field current is delayed by the direct-axis damper time constant. Any transients of field current produce also a voltage v_d in direct axis, again delayed by the direct-axis damper time constant. The damper leakage time constant, $\lambda_{kd}T_{kd}$, produces a lead effect which aids the response in both axes; see Appendix II. The field voltage produces a field current determined by the field resistance and the time constant of field- and direct-axis damper. The response of field current to a change in field voltage is aided by the damper time constant which produces a lead effect. The similarity of the equations in direct and quadrature axis is remarkable. While steady-state field current produces a quadrature-axis voltage, field current transients produce a direct-axis voltage. The damper time constants govern the response in both cases. It can be shown that the same similarity exists throughout the machine.

When load is applied to the generator, the complete signal flow diagram has to be considered. The current i_d will reduce v_q by the voltage drop i_dx_d across the synchronous reactance during steady-state conditions, as is well known. But the signal flow diagram indicates that only the voltage drop across the leakage reactance is instantaneous while the voltage drop across the magnetizing reactance x_{md} is determined by the damper time constant and leakage time constant. Similar conditions exist for the effect of i_q on v_d, only that the quadrature reactance x_q and quadrature damper time constant are involved.

The effect of i_d on v_d is very similar to the effect of i_d on v_q, except that the reactances are replaced by corresponding inductances and the voltage is affected only during current transients. Naturally, the armature resistance will appear in connection with armature leakage inductance. Again, response is affected by the damper time constant. The same applies for the effect of i_q on v_q. The interchange of reactance and inductance in the different reference axes had been noted already in the field circuit.

Any transient in direct-axis current i_d will also produce a field current i_f. This term is represented by an induced voltage (pi_dx_{md}) which generates a current in the field circuit. A more striking similarity to the field transmittance can be seen if the transmittance from i_d to i_f is rewritten as follows:

$$\frac{1}{r_f}\frac{1+pT_{kd}}{1+p(T_{do}+T_{kd})+p^2T_{do}T_{kd}\times(\lambda_f+\lambda_{kd}-\lambda_f\lambda_{kd})}\times\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}px_{md}$$

The last term represents the induced voltage with the damper winding time constants, a term familiar from other parts of the signal flow diagram. The remaining terms represent the field circuit as described by the transmittance from v_d to i_f. The voltage induced by i_d acts, therefore, on the same field circuit as the field voltage.

These equations are much too complicated to determine transients resulting from arbitrary load changes except by means of a computer. The problem is greatly simplified if all damper time constants and armature resistance are neglected and if inductive load ($R_L=0$) is considered.

$$v_d=-x_dpi_d+x_qi_q+x_{md}pi_f$$

$$v_q=-x_di_d-x_qpi_q+x_{md}i_f$$

$$v_f=-x_{md}pi_d+(1+pT_{do})r_fi_f$$

$$i_d=1/x_Lv_q$$

$$i_q=-1/x_Lv_d \qquad (13)$$

The signal flow diagram, Fig. 3, shows that for inductive load ($R_L=0$), the two transmittances between direct- and quadrature-axis voltage and current disappear, leaving only transient terms and armature resistance. Armature resistance will still produce coupling. But, also, transient terms exist proportional to direct and quadrature reactance. Changes in field current will produce a voltage in direct axis. Contrary to the simplified equivalent circuit of a synchronous generator with inductive load, which is similar to a transformer, it can be noted here that a direct-axis voltage *and* quadrature-axis

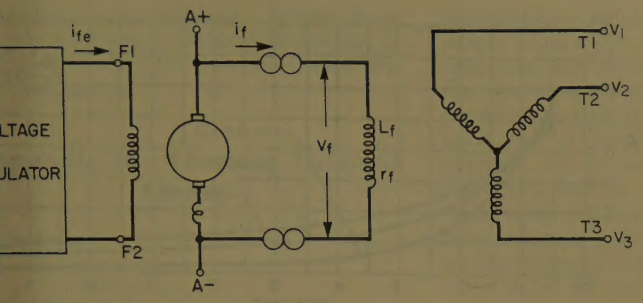


Fig. 4. Rotating exciter schematic

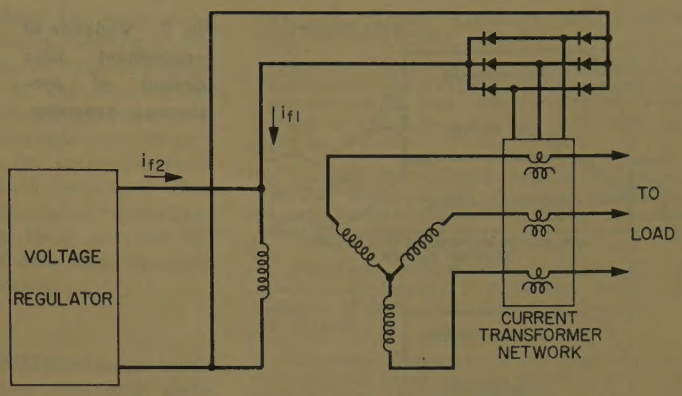


Fig. 5. Static exciter schematic

ent exist during any transient condi-
. Even load transient will produce a
sient in field current which, in turn,
luces a voltage in direct axis.

thods of Field Excitation

he previous section reviewed the gen-
equations for the transient behavior
he synchronous generator with bal-
ed load. In order to determine the
sient behavior of the terminal volt-
the source which provides the field
age v_f must be considered.
here are basically three different
hods of excitation:

Use of an integral d-c generator with
t field control.

Use of an integral a-c exciter with rotat-
rectifiers which supply the generator
resulting in a completely brushless sys-

Use of a static exciter consisting of cur-
transformers, rectifiers, and some sort
ltage regulator which receives its input
er from both generator output voltage
line current.

or generators rated 10 kva and lower,
mber of other excitation schemes be-
e practical, such as permanent-mag-
field structures with saturable back
in the stator to enable voltage regula-

he choice of one method of excitation
r any other depends on a number of
ors including cost, space available for
allation of the generator, field main-
nce, performance under fault condi-
s, and transient response. Obviously,
single method will yield the optimum
gn for all installations. At present
the transient performance of the
ous methods will be considered.
he best way to handle transient anal-
when a voltage regulator and exciter
involved is to set up equations 11 and
n an analog computer and throw
ches representing load change, field
age changes, etc. This does not
a physical understanding of the
blem, but in order to choose among the
ous excitation schemes one must be
able to evaluate the ultimate capabilities of each scheme without having to sift through a large volume of computer traces.

To do this, equation 11 will be examined in more detail, and the effect of each type of excitation scheme on v_f and i_f will be determined.

Use of a rotating d-c exciter is shown schematically in Fig. 4. Changes in generator voltage due to changing exciter output can be obtained from the equations by eliminating i_d and i_q, provided load is constant. The procedure is straightforward and results in a transfer function between v_q and v_f which is basically a single lag with the time constant T_{do}.

The more important transient is caused by load switching. It has been stated that simultaneous solution of equations 5 and 11 is much too complicated even for prescribed variations in v_f. Significant results, however, can be obtained from the simplified equation 13 if it is assumed that the exciter output voltage, which is also the generator field voltage v_f, does not change during the transient. This is a reasonable assumption because the magnetic energy stored in the exciter cannot change suddenly, hence v_f can change no faster than the exciter time constant will permit. A second assumption is that the change in load only affects i_d, with i_q remaining constant. By considering all currents and voltages in equati on 13 as incremental quantities,* terms in v_f and i_q may be dropped and after simplification the following is obtained:

$$v_d = -\frac{x_d + x'_d T_{do} p}{1 + T_{do} p} p i_d \qquad (14)$$

$$v_q = -\frac{x_d + x'_d T_{do} p}{1 + T_{do} p} i_d \qquad (15)$$

This result was obtained by Klokow.[5]

* The correct procedure would be to perform Laplace transformations using proper initial conditions. The procedure used here is a short cut whose validity rests on proper interpretation of results.

Superimposed on this transient is a second transient due to a subsequent change in v_f with i_d held constant. This is obtained from equation 13 as

$$v_d = \frac{T_{fd}' p}{1 + T_{do} p} v_f$$

$$v_q = \frac{T_{fd}'}{1 + T_{do} p} v_f$$

Since v_f itself has the form $v_f/(1+T_e p)$ (where T_e is the exciter time constant), the output voltage transient involves a double lag. Actual transients for application and removal of load also show that the response consists of one transient which dies out at the rate T_{do} and another which depends on exciter response T_e.

The basic characteristic of a system with a-c exciter and rotating rectifier can be no different. Forcing cannot change the transient unless additional dissipation is provided because the problem is basically one of energy storage in the exciter.

A significant change in transient response is possible by means of compounding as in the third method of excitation. One form of such a scheme is shown schematically in Fig. 5. As shown, the generator field current is given by

$$i_f = i_{f1} + i_{f2}$$

In order to simplify the mathematics it will be assumed that the current transformer network is so designed that $i_{f1} = a i_d$. Other compounding schemes such that i_{f1} is proportional to load current i_L or to the reactive component of i_L only are practical and yield similar results.

Using $i_f = a i_d + i_{f2}$ in equation 13 gives

$$v_d = -(x_d - a x_{md}) p i_d + x_q i_q + x_{md} p i_{f2}$$

$$v_q = -(x_d - a x_{md}) i_d - x_q p i_q + x_{md} i_{f2}$$

$$v_f = [a r_f - (x_{md} - a x_{ff}) p] i_d + r_f (1 + T_{do} p) i_{f2}$$

If one chooses $a \approx 1$, then $i_{f2} << i_f$ and it becomes practical to supply i_{f2} from a high-impedance source so that i_{f2} is substantially independent of v_f. To a first approximation at least i_{f1} will be constant

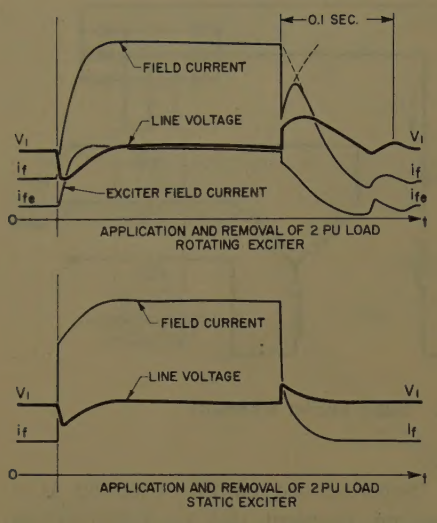


Fig. 6. Comparison of transient response of 20-kva generator with different exciters

during the load current transient and, neglecting changes in i_q, the voltage change due to a change in i_d is given simply by

$$v_d = -(x_d - ax_{md})pi_d$$

$$v_q = -(x_d - ax_{md})i_d$$

which can be made arbitrarily small by choosing ax_{md} nearly equal to x_d. Any further voltage transient then is due to a change in i_{f2}. Thus, the basic advantage derived from compounding is a much better transient response than obtainable when the generator field is supplied from a low-impedance source such as a rotating exciter. This is due to the fact that the effective synchronous reactance can be made small so that the controlling portion i_{f2} of the field current becomes very small. It is then possible to furnish i_{f2} from a fast-current source rather than a voltage source which contains stored energy and hence must have a time delay.

A comparison of the transient response of a 20-kva generator with different exciters is given in Fig. 6.

Appendix I. Transfer Function of Generator

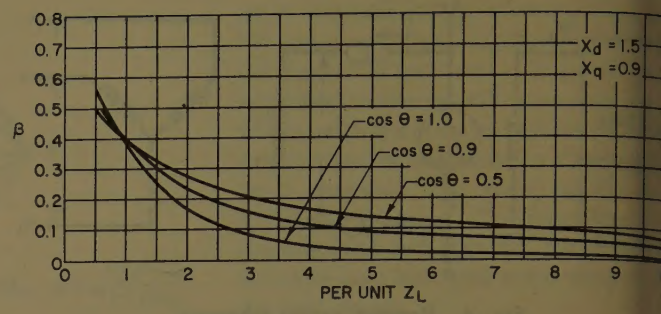
To facilitate the derivation of a transfer function and to demonstrate the effect of machine parameters and load on gain and time constant of the generator, the machine is assumed to be linear. The main field in the d-axis accounts for the predominant time constant. The effect of damper windings shall be neglected. The objective is to determine v_q, v_d, and v_t as function of v_f.

The load representation is assumed not to be influenced by a change in frequency. The following equations apply:

$$v_d = r_L i_d - i_q x_L$$

$$v_q = r_L i_q + i_d x_L \qquad (16)$$

Fig. 7. Variation in predominant time constant of synchronous generator



From equations 1, 2, and 3 there follows, when armature resistance and unidirectional flux transients are neglected:

$$v_q = x_{md} i_f - (x_{md} + x_{al}) i_d$$

$$v_d = (x_{mq} + x_{al}) i_q$$

$$v_f = i_f r_f + p i_f (x_{md} + x_{fl}) - p i_d x_{md} \qquad (17)$$

from which one obtains the field current and the load voltages:

$$i_f = \frac{v_f + p i_d x_{md}}{r_f + p(x_{md} + x_{fl})}$$

and

$$v_q = \frac{v_f}{r_f} x_{md} \left[1 - \frac{x_d}{x_d + x_L + \dfrac{r_L^2}{x_q + x_L}} \right] \frac{1}{1 + pT_d}$$

$$v_d = v_q \frac{x_q r_L}{x_L(x_q + x_L) + r_L^2} \qquad (18)$$

with the time constant

$$T_d = \frac{x_{md} + x_{fl}}{\omega_0 r_f} \left[1 - \frac{x_{md}}{1 + \dfrac{x_{fl}}{x_{md}}} \frac{1}{x_d + x_L + \dfrac{r_L^2}{x_q + x_L}} \right] \qquad (19)$$

The terminal voltage v_t is found as the vector sum of v_d and v_q.

$$v_t = \sqrt{v_d^2 + v_q^2} = \frac{v_f}{r_f} x_{md} \times \frac{\sqrt{1 + \dfrac{x_q^2}{|Z_L|^2} + 2x_q \dfrac{\sin\theta}{|Z_L|}}}{1 + \dfrac{\sin\theta}{|Z_L|}(x_d + x_q) + \dfrac{x_q x_d}{|Z_L|^2}} \frac{1}{1 + pT_d} \qquad (20)$$

in which

$$|Z_L| \sin\theta = x_L$$

$$|Z_L| \cos\theta = r_L$$

$$|Z_L|^2 = r_L^2 + x_L^2$$

$\cos\theta$ = load power factor

Equation 20 represents in its first term, $v_f/r_f\, x_{md} = i_f x_{md}$, the steady-state no-load gain of the machine corresponding to t air-gap line. The second term accounts f the change in gain due to load. The thi term represents the dynamic characteri tics of the machine expressed by a tir constant T_d as represented by the ma field. The time constant is subject changes for different loads and differe power factors. The first term of equati 19 equals the no-load time constant T_{do}.

$$T_d = T_{do} \left[1 - \frac{x_{md}}{1 + \dfrac{x_{fl}}{x_{md}}} \times \frac{1}{\dfrac{Z_L^2 \cos^2\theta}{|Z_L| \sin\theta + x_q} + |Z_L| \sin\theta + x_d} \right] \qquad (2$$

or if the term dependent on load is denot β, the time constant of the generator equal to

$$T_d = T_{do} \left[1 - \frac{x_{md}}{1 + \dfrac{x_{fl}}{x_{md}}} \beta \right] \qquad (2$$

The factor β of equation 22 accounts for t effect of machine parameters, load, a power factor. Values of β for differe types of load are given in the curves Fig. 7.

Saturation will result in a reduced ga and a smaller time constant. The effect saturation can be accounted for by assumi that the direct-axis mutual reactance x and the synchronous reactance x_d vary wi flux level. The degree of saturation can expressed by a multiplicative saturati factor γ. The effective mutual reactan becomes γx_{md} where $\gamma = 1$ for the unsat rated machine and $\gamma > 1$ for the saturat machine. γ can be determined as the rat of the slope of the no-load saturation cur to the slope of the air-gap line obtained an operating point corresponding to ψ_m Equations 20 and 22 are still valid wh x_{md} and x_d are replaced by γx_{md} and γx respectively.

Fig. 7 shows that the time constant of synchronous generator with $x_d = 1.5$ a $x_q = 0.9$ per unit changes from T_{do} at

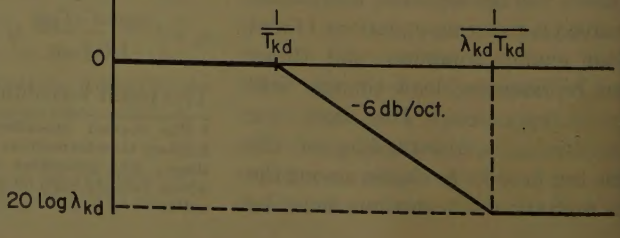


Fig. 8. Transfer function graph

to 0.6 T_{d0} at rated load. At rated load, time constant is almost independent of power factor for these particular ma- e reactances.

ppendix II. Transfer Function Graph

he term

$$H_{kd}=\frac{1+p\lambda_{kd}T_{kd}}{1+pT_{kd}}$$

ears throughout the signal flow diagram. implications of this term on the re- ise of the machine have to be fully un- tood. Considering the term as a trans- fer function which defines the ratio of an output to an input variable, a graph of this ratio e_o/e_i can be made in form of a Bode plot as shown in Fig. 8. At a corner frequency equal to the reciprocal time constant, the frequency response falls off at a slope of −6 decibels per octave, until it reaches a second corner frequency at a value equal to the reciprocal of the leakage time constant. From there on, the frequency response is flat with an attenuation equal to the per-cent leakage.

References

1. Two-Reaction Theory of Synchronous Machines. Generalized Method of Analysis—Part I, **R. H. Park.** *AIEE Transactions*, vol. 48, July 1929, pp. 716–30.

2. Analog Computer Representations of Synchronous Generators in Voltage-Regulation Studies, **M. Riaz.** *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 75, Dec. 1956, pp. 1178–84.

3. Control System Synthesis (book), **John C. Truxal.** "Electrical and Electronic Engineering Series," McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

4. Feedback Theory—Some Properties of Signal Flow Graphs, **S. J. Mason.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., pp. 1144–56.

5. Maximum Overvoltage on Aircraft A-C Generators After Sudden Removal of Load, **R. E. Klokow.** *AIEE Transactions*, pt II (*Applications and Industry*), vol. 74, 1955 (Jan. 1956 section), pp. 417–21.

Discussion

hard T. Smith (The University of Texas, tin, Tex.): The use of signal flow graphs vides another method for depicting the tionships among the variables of syn- nous machine equations. Other tech- ues consist of elimination of axes[1] and flux-linkage equivalent circuits.[2] By ns of the equivalent circuits one can te equations 10 of the paper by inspec- and thus get a quick check on the signal graph results. It is noted that equa- s 11 do not include the effects of arma- e resistance.

he authors point out the "remarkable ilarity" of the direct- and quadrature- equations at no load. Actually, even at load the equations may be greatly dis- ilar if the variables are expressed as licit functions of time, depending upon nature of i_f. It is not clear how the hors are defining reactance and induct- e in their discussion of machine charac- stics. Since the equations are in per- t form it would not seem to be significant ether the word inductance or reactance sed.

he determination of transient response arbitrary changes in load, for balanced ditions, has been investigated in an ap- ximate fashion by Concordia and others, ticularly for fixed excitation. However, thods for including variable excitation e been given. For many applications se approximate methods are satisfactory, ticularly when the original equations are ject to the inaccuracies of omitting satu- on, lumping damper effects, and doubt cerning the values of "constants."

FERENCES

A Short Course in Tensor Analysis for ctrical Engineers (book), **G. Kron.** John ey & Sons, Inc., New York, N. Y., 1942.

Starting Performance of Salient-Pole chronous Motors, **T. M. Linville.** *AIEE nsactions*, vol. 49, Apr. 1930, pp. 531–47.

o-nan Yu (Villanova University, Vil- ova, Pa.): The authors are to be con- tulated on their contribution to this d. The complexity of the relations of a chronous machine has been reduced ough the use of signal flow graphs and e constants. The paper is excellent o for educational purpose.

It is noticed that reciprocal mutual reactances, x_{md} and x_{mq}, are used in the formulation. Basically, the mutual induction between the stator 3-phase windings, *a*, *b*, and *c*, and the rotor single-phase windings, *f*, *kd*, or *kq*, is not reciprocal because the rotor reactances are due to 3-phase excitation, and the stator reactances are due to single-phase excitation. However, the reciprocity can be obtained through transformations.

It is also noticed that the authors follow Park's original work in using $-i_d$ and $-i_q$ in their formulations, whereas Kron, White, Woodson, and others use $+i_d$ and $+i_q$ in their formulations.

I am also very glad to see the authors applying signal flow graphs to the analysis of synchronous machines. By coincidence, I have presented a conference paper, "Signal Flow Graphs and Equivalent Circuits of Asynchronous Motors," which discusses the use of these graphs in the analysis of asynchronous machines. For those who still doubt the usefulness of a signal flow graph, I would point out the following facts:

1. It integrates and co-ordinates all the piecemeal information of fundamental equations into one clear picture, retaining every detail, and further, it can be easily reduced.
2. It exhibits explicitly the feedback and the feedforward loops of a system and suggests means of compensation and control.
3. It can be used to describe an electrodynamic system, including inertia, damping, speed, and torque, in addition to static relations.

In my paper, signal flow graphs of the Schrage motor and double-fed induction

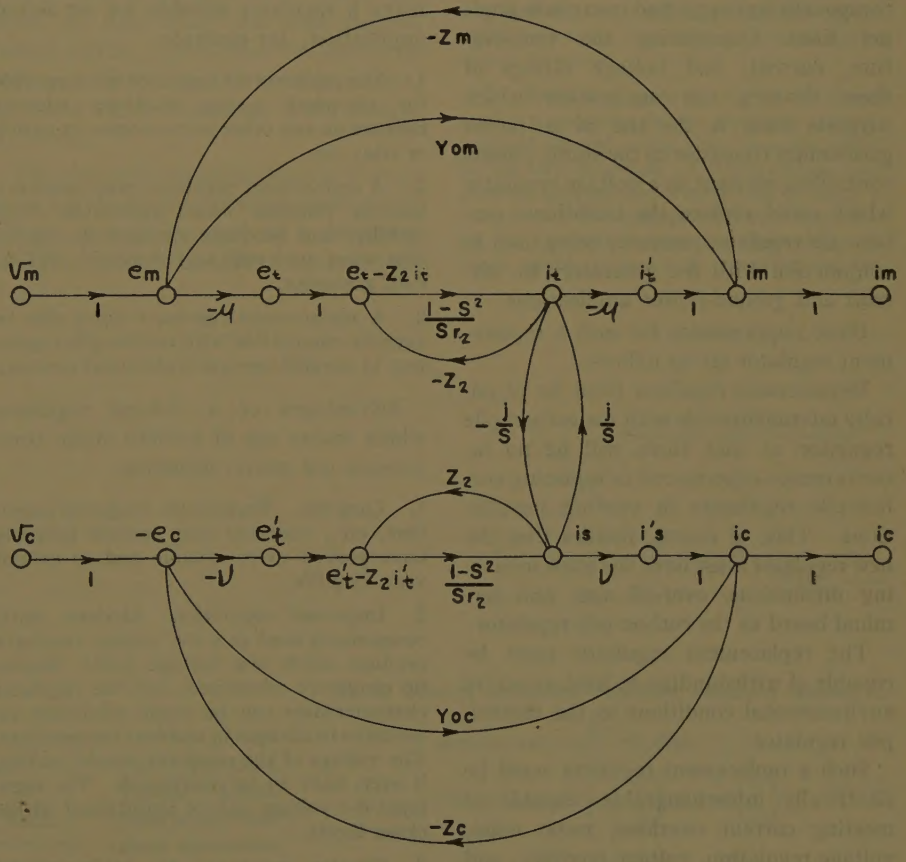


Fig. 9. Signal flow graph of 2-phase servomotor

motor by revolving-field theory, and of the single-phase induction motor and 2-phase servomotor by cross-field theory, are shown. Equivalent circuits are derived from the signal flow graphs.

For the interest of the readers, one of the signal flow graphs is shown in Fig. 9.

A. Krausz and **H. A Kahle**: We gratefully acknowledge the discussions by R. T. Smith and Yao-nan Yu.

In this paper, an attempt was made to use the techniques of feedback control systems to describe synchronous generators in a form suitable for synthesis of systems which involve control of synchronous generators, and where transient behavior is of primary interest. For analysis of the steady-state characteristics or for design of generators, conventional methods are to be preferred.

The similarity between direct- and quadrature-axis equations at no load as given by equation 12 applies to the response of quadrature- and direct-axis voltage to field current and rate of change of field current, respectively.

All equations except 1 and 2 are written in per-unit form; hence, all quantities are by definition nondimensional. To obtain dimensional equations, all per-unit quantities in the per-unit equations must be mu plied by their corresponding base quanti which have dimensions. Since in the p unit equations, the base quantities for x_{md}, etc., have the dimension of reacta (ohms), not inductance, the symbol x rat than L is used, and the x quantities properly referred to as per-unit reactanc

The signal flow graph is one of ma analytical tools which, when properly plied and interpreted, can become use for synthesis and compensation of com cated control systems. It is hoped that st highly developed techniques of control s tem synthesis will be found increasingly u ful in the field of electric power control.

A Transistorized D-C Voltage Regulator for Direct Replacement of Carbon-Pile Regulators

P. D. COREY
NONMEMBER AIEE

W. O. HANSEN
NONMEMBER AIEE

FOR SEVERAL years germanium transistors have been available which are capable of controlling appreciable power. The availability of this versatile component has suggested many new product lines. Considering the temperature, current, and voltage ratings of these devices, one application which suggests itself is the use of a power germanium transistor as the static, power-controlling element in a voltage regulator which could replace the traditional carbon-pile regulator presently being used in conjunction with d-c generators in aircraft and ground-power applications.

Basic requirements for such a replacement regulator are as follows:

Replacement regulator must be physically interchangeable with the carbon-pile regulator so that there will be no inconvenience experienced in replacing carbon-pile regulators in existing installations. This, of course, implies that the new regulator must have the same mounting dimensions, over-all size, and terminal board as the carbon-pile regulator.

The replacement regulator must be capable of withstanding at least as severe environmental conditions as the carbon-pile regulator.

Such a replacement regulator must be electrically interchangeable, capable of meeting current overload, radio noise, voltage regulation, voltage recovery, and other similar electrical requirements.

With respect to electrical interchangeability, *MIL-R-6809* describes the carbon-pile regulator but does little to really define electrical performance required to make a regulator suitable for an actual application; for example:

1. The replacement regulator must provide for electrical system build-up without reliance on any other power source or switch or relay, etc.

2. A replacement regulator must possess a transfer function which guarantees both stability and adequate steady-state regulation when used with any standard *MIL-G-6162* generator.

3. A replacement regulator must also be entirely compatible with carbon-pile regulators in parallel operation electrical systems.

Advantages of a voltage regulator which makes use of modern static components are many, including:

1. Long life. Transistors, magnetic amplifiers, etc., operated and applied properly have proved to be reliable and to exhibit very long life.

2. Improved regulation. Modern static components used in a d-c voltage regulator produce much less voltage drift. Warm-up errors are minimized, and the regulator characteristics can be made relatively insensitive to changes in ambient temperature. The voltage of the regulator should seldom, if ever, have to be readjusted. The regulated d-c voltage can be maintained within closer limits.

3. Greater resistance to environmental conditions. The static d-c voltage regulator can withstand the vibration encountered an aircraft without having to use sh mounts. The static d-c voltage regula is not affected by the position of mounti or by linear acceleration.

4. Improved speed of response. The sta components used in the transistorized regulator are capable of operating faster th are components which must physically mo to produce the desired electrical change.

5. Easier to apply. The static, transist ized d-c voltage regulator can be bet represented mathematically as an element a servosystem; thus, expected performan can be more accurately predicted.

The regulator circuit and product d sign to be described is only one of ma possible approaches to arrive at a tra sistorized regulator for d-c generato however, this approach appears best f aircraft and ground-power applicatio Advantage is taken of the latest prov components in a straightforward circ configuration in which reliability is pa ticularly emphasized. Tests have prov that this regulator not only possesses t advantages inherent in a static elect device but that all practical system r quirements are adequately met.

Basic Principle of Operation

Fig. 1 illustrates how a carbon-p regulator is used to control a self-excit shunt field d-c generator. Briefly, t carbon stack is a pressure-controll variable resistance element which is us to control generator field current in t same manner as one would employ a fi rheostat. While there is certain nothing wrong with this principle operation (except possibly the nec

Paper 60-195, recommended by the AIEE A Space Transportation Committee and approved the AIEE Technical Operations Department presentation at the AIEE Winter General Meeti New York, N. Y., January 31–February 5, 19 Manuscript submitted March 24, 1959; made av able for printing January 18, 1960.

P. D. COREY and W. O. HANSEN are both w General Electric Company, Waynesboro, Va.

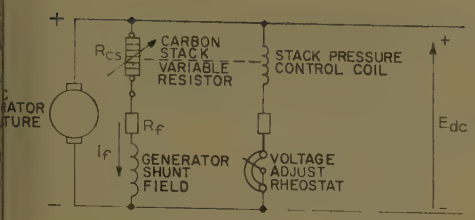
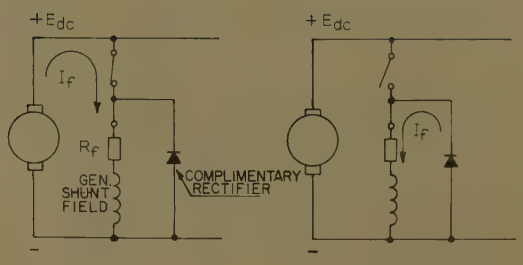


Fig. 1 (left). Generator field current control utilizing a variable series field resistance element



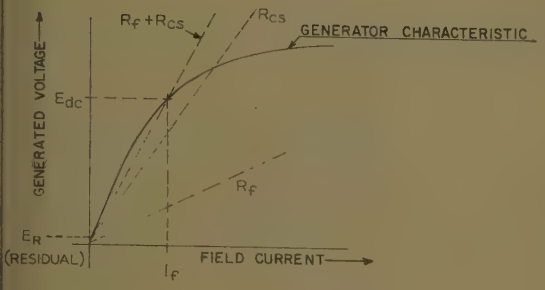
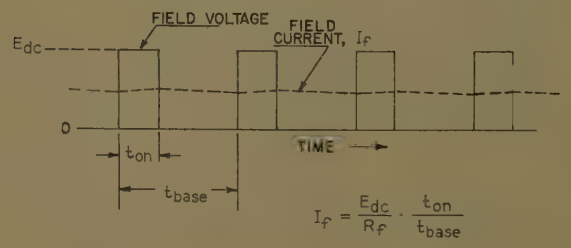


Fig. 2 (right). Generator field current control utilizing an oscillating switching element

$$I_f = \frac{E_{dc}}{R_f} \cdot \frac{t_{on}}{t_{base}}$$



y high power loss in the carbon-stack tor), the carbon stack and the pres- connections to it are inherently sub- to wear from mechanical movement electrical arcing. Also, vibration, r accelerations, and extreme environ- tal conditions may adversely affect performance of the carbon-pile reg- or.

one attempts to utilize an available anium power transistor as a variable- tance element, one finds the power pation to be too high to permit +71 C rees centigrade) ambient operation, if a number of transistors are lleled. Thus, if one wishes to utilize rmanium power transistor for field rol, some other operating principle t be sought.

g. 2 shows how a switch oscillating relatively high frequency could be to control field current of a shunt d-c generator. During the ON interval, field voltage is equal to generator armature voltage and field current flows through the switch. During the OFF interval, field voltage is nearly equal to zero and field current goes through the "complementary" rectifier. If the repetition rate of the switching operation is high enough, the inductive field will maintain very smooth current, and negligible modulation of output voltage of the generator will occur. Field current will be merely the average field voltage divided by field resistance. It is evident that control of the ON and OFF times will result in controlling the field current (which will in turn control the generated voltage). If the switch used is perfect, no power will be dissipated in the switching device.

The preceding suggests that if a transistor has the required voltage and current ratings, it may be used as a switching device to control the field of a d-c aircraft generator without the power dissipation which is incurred in a variable resistance control device. Practical limitations, however, do require appreciable power dissipation in a germanium power transistor even if it is used as a switch. This dissipation stems from three principal causes:

1. Referring to Fig. 3, it is seen from the common emitter characteristic shown that a typical germanium power transistor does exhibit an appreciable saturation resistance when turned full ON. A typical value would be 0.05 ohm,. At the field currents required in practical applications, appreciable power is, therefore, dissipated during the ON interval.

2. Referring again to Fig. 3, it is seen that a transistor exhibits a certain collector leakage current, even when heavily biased in the OFF direction. Circuit design must provide appreciable reverse bias in order to assure that this leakage current is very close to the absolute minimum achievable. At room ambient junction temperature, this

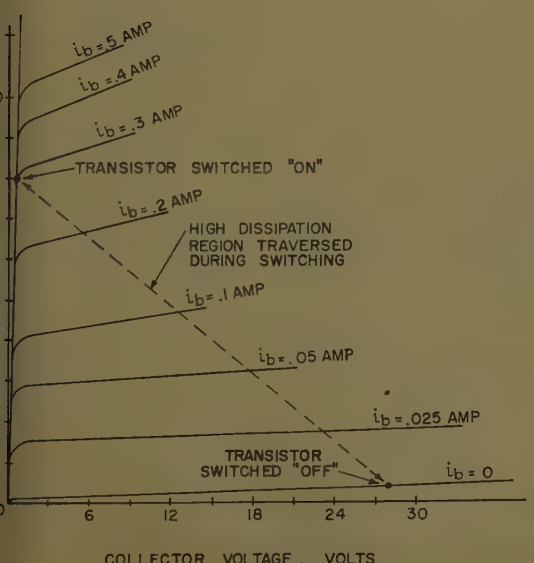


Fig. 3 (left). Typical germanium power transistor characteristics, common emitter

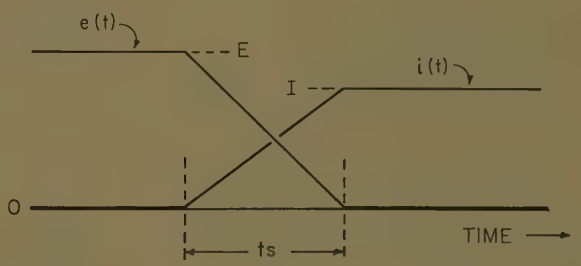


Fig. 4. Energy dissipation during the switching interval, t_s.

$$\text{Watt-seconds} = \int_0^{t_s} e(t)i(t)dt$$

$$\text{Where: } e(t) = E\left(1 - \frac{t}{t_s}\right) \quad i(t) = I\left(\frac{t}{t_s}\right)$$

$$\text{Thus, } W = \int_0^{t_s} \left(\frac{EIt}{t_s} - \frac{EIt^2}{t_s^2}\right)dt = \frac{EIt_s}{6}$$

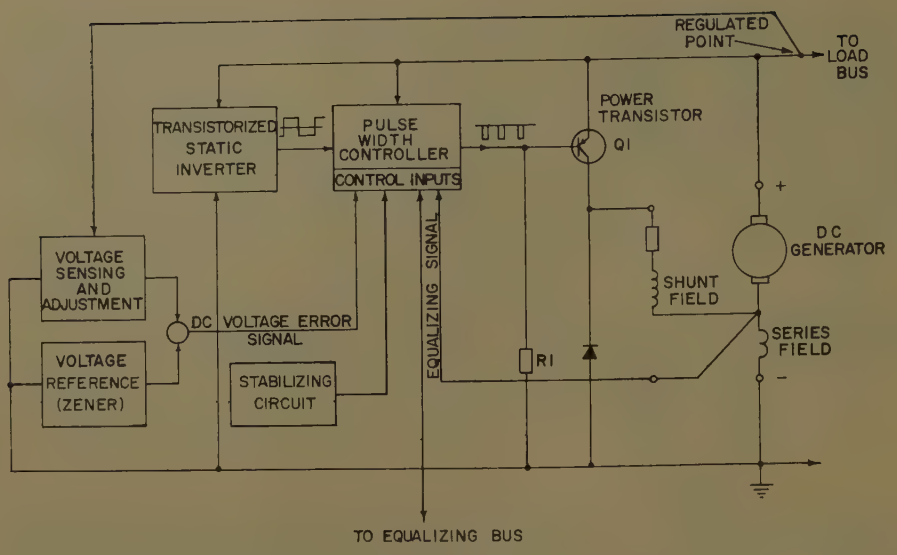


Fig. 5. Block diagram; basic configuration of transistorized regulator

leakage current is negligible, and results in very minute additional junction dissipation; however, at elevated junction temperatures which are encountered at elevated ambient temperatures, one must seriously consider power dissipation from this source. It is found that with proper circuit design, the higher dissipation results from the collector saturation resistance rather than from leakage current even at high ambient and resulting high junction temperatures. Thus, maximum dissipation occurs at maixmum load current.

3. The largest source of dissipation occurs in the switching interval when the transistor is being switched from its blocking to its conducting state (full OFF to full ON) and vice versa. Fig. 4 shows the voltage-current-time relationships assuming linear switching. On the attached derivation, it is seen that the expression for the watt-second loss per switching operation is

$$W = \frac{EIt_s}{6}$$

where

E = d-c supply voltage in volts
I = load current in amperes
t_s = switching interval in seconds

Transistor dissipation for a typical set of "worst condition" values is as follows:

$P_s = fw = 2.98$ watts where $f = 4$ kc, $E_{dc} = 28$ volts, $I = 8$ amperes, $t_s = 20 \times 10^{-6}$ seconds

$P_{RS} = i^2 R_s \left(\frac{T_{on}}{t_{base}}\right) = 1.6$ watts where saturation resistance = 0.05, $t_{on}/t_{base} = 0.5$

Total = 4.58 watts

where:

t_{on} = ON time interval
t_{base} = reciprocal of frequency.

As can be seen from these data, a transistor having a 5-C-per-watt junction rise above ambient can control 8 amperes continuously at +71 C and stay within the temperature rating of the device.

Circuit Configuration

Once it was decided to use a power transistor operated in its switching mode to control generator field current, considerable investigation was done to decide what switching circuit configuration would be best suited to this particular application. The following are some of the factors which were considered in addition to the requirements already noted in the introduction.

1. Reliability: The circuit must be of such a nature as to cause the most reliable switching of the output transistor. Wave fronts should be steep, and forward current and back bias values must be more tha adequate to assure minimum transist dissipation and a considerable margin safety from "thermal runaway" at hig temperature.

2. Control Range: In order to assure th best in transient performance, it is desired be able to turn the output transistor from condition of completely OFF 100 per cent the time to an almost fully ON condition.

3. Stability: To minimize voltage drif it is desirable to achieve the pulse widt modulation control with a device which relatively insensitive to drift caused b aging and ambient temperature changes.

4. Electrical Isolation: If compatibilit with the present carbon-pile system is to b achieved, electrical isolation is required b tween the voltage error sensing circuit a the equalizing circuit.

Fig. 5 is a block diagram showing th circuit configuration finally decided upo

The d-c line voltage is compared with stable voltage reference, and a sign proportional to the regulated volta error is supplied to the pulse width mod lation controller.

This controller consists of a magnet amplifier of the full-wave center-tapp configuration which is used in conjun tion with a rectified bias supply. Bo portions of this circuit receive their pow from a transistorized static invert in which switching transistors are corporated in a multivibrator oscillat The square-wave output of this inverter ideal for the pulse controller.

The magnetic amplifier used for a pul width control is exceptionally reliab and exhibits very low drift with temper ture and age. Also, the use of the ma netic amplifier automatically provid an easy means for isolating the er sensing and equalizing circuits.

The output of the controller drives t

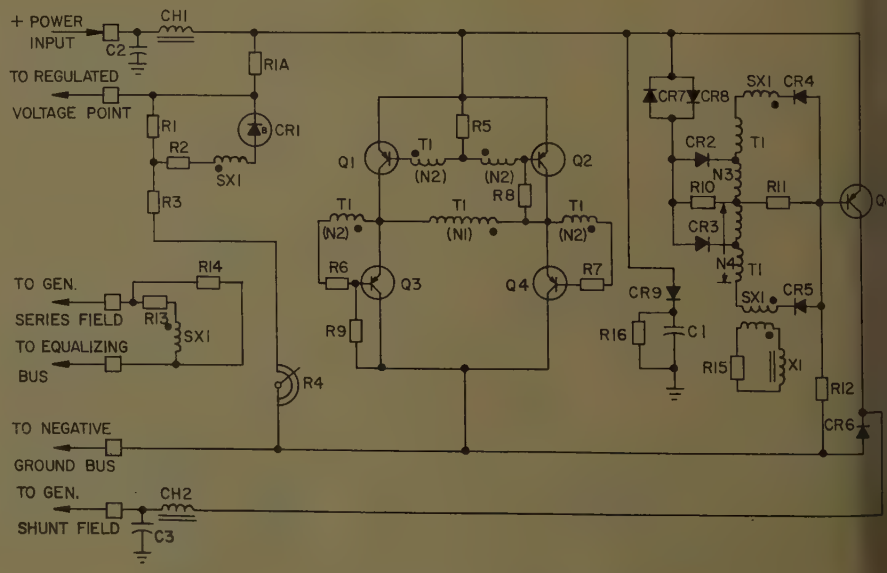


Fig. 6. Final circuit diagram of transistorized regulator

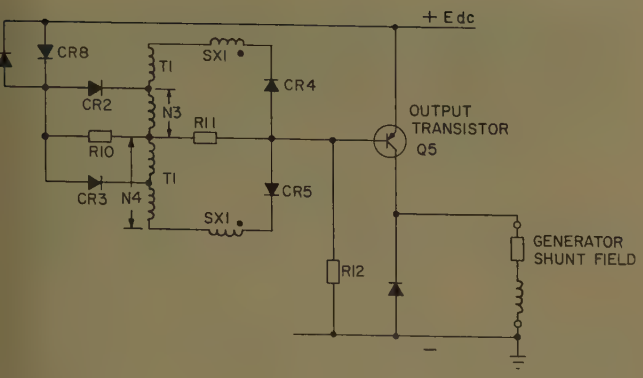


Fig. 7. Output transistor and associated pulse width modulation control

e of the power transistor used to con- field current. The "complementary ifier" provides a path for the field cur- during OFF intervals of $Q1$. $R1$ is nected from the base of $Q1$ to the nega- (–) terminal so as to bias $Q1$ ON ing initial build-up. When initial d-up has proceeded until line voltage excess of about 3 volts, the transis- zed inverter starts supplying power he pulse width modulation controller normal control of $Q1$ is obtained. us, initial build-up is reliably accom- hed without employing an external tery or power supply or mechanical tching device.

Circuit design is such that $Q1$ is always rated reliably as a switch, never as linear amplifying element. Hence, nges of $Q1$'s characteristics due to nperature or aging will have a negli- le effect on regulator operation.

In order to assure proper transient re- nse performance, stabilizing circuitry incorporated in this design. Final tem stabilization was accomplished by heoretical analysis later confirmed by oratory system testing. The circuit ign provides an adequate phase mar- to insure system stability for all in- ded operating conditions, including use of any *MIL-G-6162* generator, l provides high system gain for ob- ning close, steady-state voltage regu- on.

tailed Circuit Description

. Sensing circuitry: Referring to Fig. the sensing circuit ($R1$, $R2$, $R3$, $R4$, 1, and the control winding of $SX1$) sists of a conventional voltage error sing circuit in which a portion of the voltage is compared with the "break- vn" voltage of a silicon "Zener" diode. Zener voltage reference device was sen because of its reliability, stability, l accuracy, along with its very small and light weight. Rheostat $R4$ pro- es the desired voltage adjustment. 4 prevents the system from going to "ceiling" in the event the voltage sensing wire is inadvertently open-circuited.

2. The transistorized static inverter (consisting of $Q1$, $Q2$, $Q3$, $Q4$, $R5$, $R6$, $R7$, $R8$, $R9$, and $T1$) is a bridge-type inverter circuit. Transistors are turned on in alternate pairs, applying power to the primary of transformer $T1$ much in the same manner as a double-pole double-throw switch. Drive to transistors $Q1$, $Q2$, $Q3$, and $Q4$ is controlled by resistors $R5$, $R6$, and $R7$, in conjunction with the associated $T1$ secondary windings. "Turn on" drive is merely the line voltage times the transformer ratio divided by the series resistance:

$$i_b = \frac{E_{dc}}{R5} \cdot \frac{N_2}{N_1}$$

Since very small base current flows in the reverse direction, for all practical purposes the reverse bias is:

$$E_{rev} = E_{dc} \cdot \frac{N_2}{N_1}$$

The ON collector current for each of the transistors is given by adding the excitation current for $T1$ to the currents in the various secondaries reflected back to the primary winding.

The frequency of oscillation of this inverter is easily determined by the following:

$$E_{dc} = 4.0 B_m N_1 A_{fe} f \times 10^{-8}$$

$$f = \frac{E_{dc} \times 10^8}{4 B_m N_1 A_{fe}}$$

where

E_{dc} = d-c supply volts
N_1 = number of primary turns
B_m = maximum core flux density in lines per square inch
A_{fe} = core area

The size of this device, particularly the transformer $T1$, will be dependent upon the nominal frequency of operation chosen—the higher the frequency, the smaller the transformer. However, it will be remembered that the switching transient accounts for a major part of the watts dissipated at the power output transistor junction; hence, it is evident that the maximum frequency permissible will be dictated by the output transistor dissipation allowable.

3. Pulse width modulation control: Fig. 7 shows the portion of the regulator circuit involved in generating and controlling the drive pulses for the output transistor.

During initial build-up current flows from the positive bus, through the emitter-base junction of $Q5$ and through $R12$ to the minus terminal of the generator. This turns transistor $Q5$ full ON, which in turn supplies full excitation to the generator shunt field. Build-up proceeds in this fashion until there is approximately

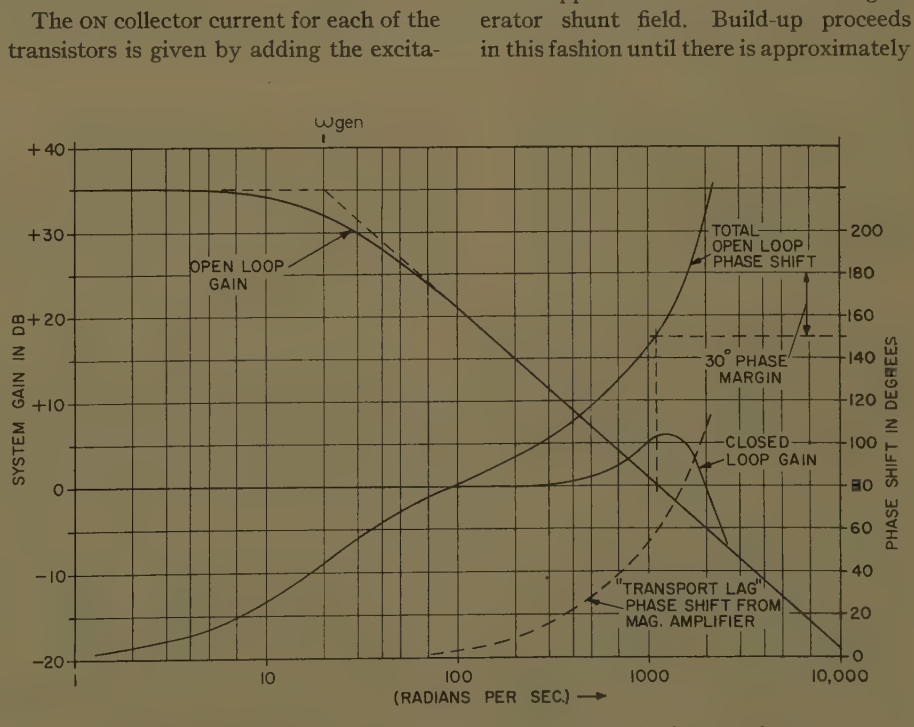


Fig. 8. Bode diagram of generator-regulator system with "fast" regulator

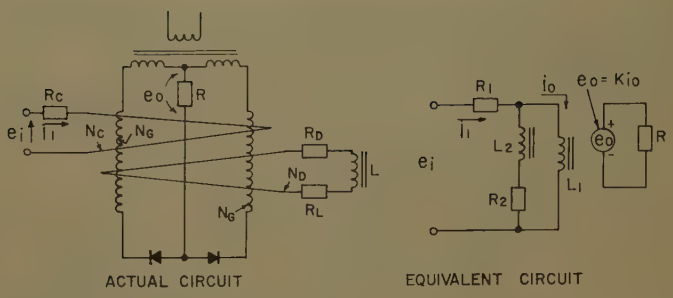


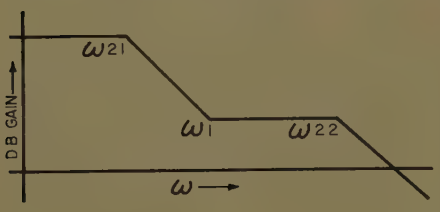
Fig. 9. Modification of amplistat transfer function to provide stabilizing

From the equivalent circuit:

$$\frac{e_o}{e_i}=\frac{K}{R1}\cdot\frac{s\frac{L_2}{R_2}+1}{s^2a+sb+1}$$

where $a=\frac{L1L2}{R1R2}$ and $b=\frac{L2}{R2}+\frac{L1}{R2}+\frac{L1}{R1}$

Since the "damping ratio" of the quadratic lag will exceed unity, the following Bode diagram applies:



$$\omega_1=\frac{R2}{L2}\qquad \omega_{21}=\frac{2}{b+\sqrt{b^2-4a}}$$

$$\omega_{22}=\frac{b+\sqrt{b^2-4a}}{2a}$$

Equivalent circuit values:
R1 = RC

$$L_1=\frac{2N_c^2AfeBm\cdot10^{-8}}{lfe(H_{c,2}-H_{c,1})}$$

$$L_2=L\left(\frac{N_c}{N_d}\right)^2;\quad R2=(R_L+R_D)\left(\frac{N_c}{N_d}\right)^2$$

$H_{c,1}$ = dc coercive force (ampere-turns per inch)
$H_{c,2}$ = ac coercive force (ampere-turns per inch)
Bm = maximum flux density
Afe = core area (square inches)
lfe = core length (inches)

3 volts output from the generator. From this point on, the transistorized static inverter starts applying a-c power to the magnetic amplifier *SX*1, and normal switching control of output transistor *Q*5 is achieved.

It should be noted in connection with initial build-up that silicon rectifiers *CR*7 and *CR*8 serve to block the flow of current through resistors *R*10 and *R*11 when the voltage across the rectifiers 7 and 8 is below the "barrier potential" of these devices (about 0.7 volt). The only function of these rectifiers is to assure that during the first portion of the initial build-up, current through resistor *R*12 will pass through transistor *Q*5, and not through the pulse generating circuitry. This feature assures positive build-up.

At normal generator output voltage, control of *Q*5 is achieved in the following manner:

During the interval when the magnetic amplifier *SX*1 is not saturated, a d-c bias voltage, provided by rectifying the output of the static inverter by components *CR*2, *CR*3, and *R*10, turns the transistor *Q*5 OFF. Sometime later during the half-cycle of the a-c inverter supply voltage, *SX*1 will saturate. The resulting d-c voltage developed across resistor *R*11 opposes the voltage of the bias supply so that base current suddenly flows in the direction to turn the output transistor ON and apply full line voltage to the generator shunt field. The portion of each half-cycle of the inverter output during which *Q*5 will be turned OFF is governed by the "firing angle" or "excitation interval" of *SX*1, which in turn depends upon the control current. Since the control winding is in the error sensing circuit as previously described, the percentage of time that transistor *Q*5 will be turned on will be a function of the line voltage as compared to the voltage reference. The circuit design is such to cause the magnetic amplifier *SX*1 cover its full range with a 1-volt excurs of the d-c line voltage.

The use of a square-wave a-c sup voltage is ideal because it allows change of "firing angle" of *SX*1 with c trol current to be almost linear. A the output pulses of the magnetic am fier are characterized by the fast rise ti and a rectangular shape, while the reve bias circuit has almost a perfect d-c vo age output, requiring no filtering.

4. Equalizing circuit: Referring ag to Fig. 7, the equalizing circuit wh enables this regulator to share load c rent properly during parallel operati consists of resistors *R*13, *R*14, and a seco control winding wound on magnetic am fier *SX*1. This circuit is designed that it is in every way externally identi with the equalizing winding of the carb pile regulator. This insures compatibil in a system where a transistorized regu tor might be used to control one genera connected in parallel with another gene tor which is controlled by a carbon-p regulator. The values of the resist and the equalizing winding are chosen provide a resistance identical to the c bon-pile regulator. The ratio of resist *R*13 and *R*14 is chosen such that fo given equalizing signal, droop in regula voltage will be identical to that exp enced when an equal signal is applied the equalizing coil of a carbon-pile regu tor. As was mentioned before, a pri advantage of using a magnetic amplif for pulse width modulation control that the equalizing signal circuit c readily be electrically isolated from t remainder of the regulator circuit.

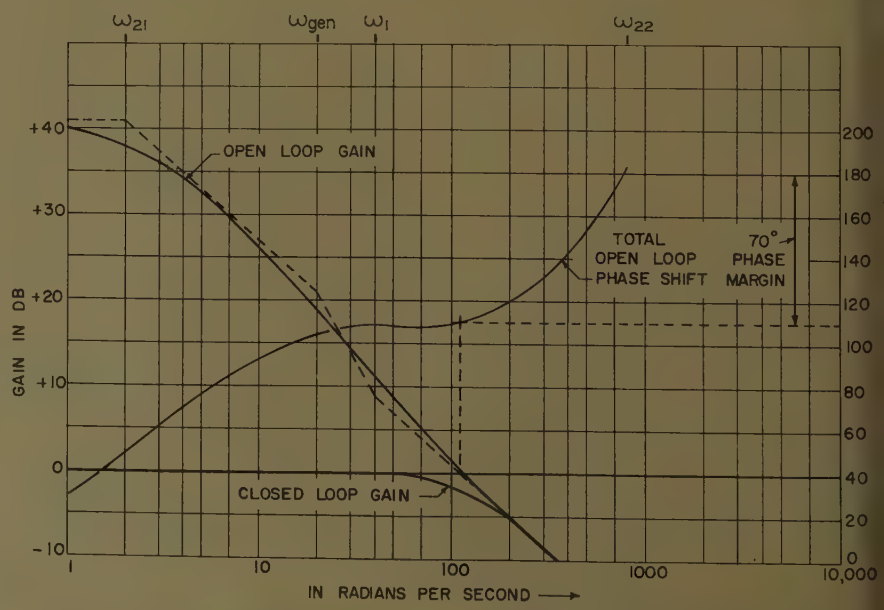


Fig. 10. Bode diagram of generator-regulator system with inductor stabilizing added

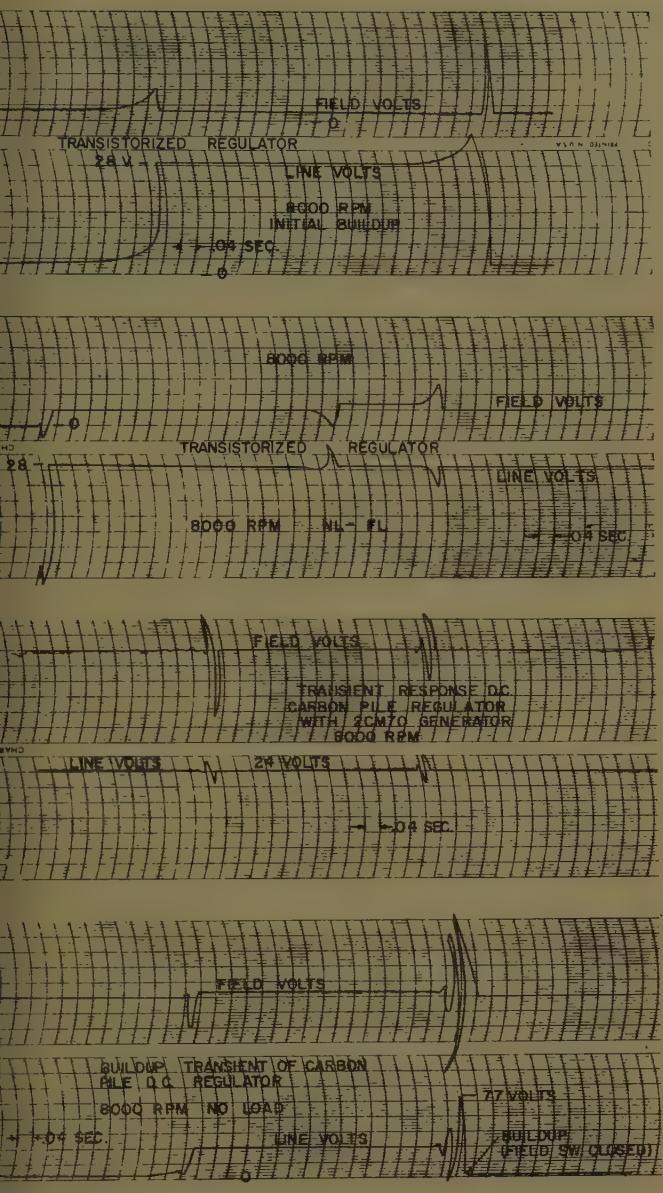


Fig. 11. Typical single generator performance of the transistorized regulator (top traces) compared to the carbon-pile regulator (bottom)

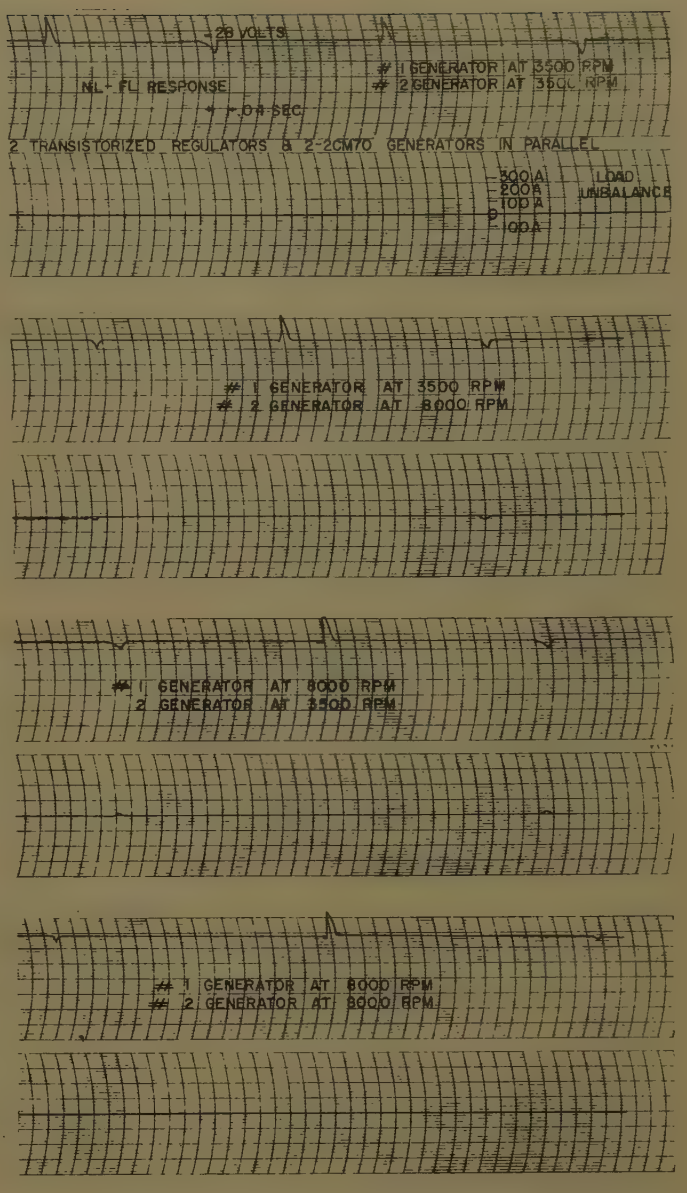


Fig. 12. Typical parallel operation of systems with transistorized regulators over the range of generator speed extremes

5. Stabilization: In order to assure ansient stability while at the same me maintaining adequate steady-state ain for close voltage regulation, it is ecessary to provide stabilizing circuitry r this regulator. This assumes a very mple form, consisting of an inductance ith a series resistor connected across a ird control winding wound on the magetic amplifier *SX*1.

During the initial design phase of this gulator, it was felt that it would be most esirable to make the regulator operate ery "fast" so that the generator field time onstant would be the only significant ource of phase shift and attenuation. It as decided to design *SX*1 and the associted voltage error sensing circuit so that alf-cycle response time was achieved. In his case, a time constant does not apply o the transfer function of the magnetic mplifier; instead, the magnetic amplifier is representable as a "transporation lag" with increasing phase shift as a function of input signal frequency, but no attenuation. Fig. 8 is a Bode diagram of the original system using the "fast" regulator design. According to this initial analysis, the system was considered to be very stable over a wide range of generator characteristics.

Later in the design phase, however, it was found that this system using the "fast" regulator was susceptible to high-frequency noise disturbances. These disturbances varied in amplitude from generator to generator, and were found to consist of two distinctive components. A band of high-frequency noise was found to exist from 1,000 radians upward as a result of commutator "hash." This noise was prevalent, particularly at no load and without a battery connected to the d-c bus. A band of lower frequency disturbances was found in the 300- to 1,000-radian-per-second range. This "noise" source, apparently caused by generator rotor eccentricity, was found to be at a frequency directly proportional to the generator rotational speed, and to be relatively independent of load on the generator bus. While these noise input signals did not result in a true instability of the regulated system, the "fast" regulator exhibited undesirable response which resulted in large components of noise on the field voltage. In some instances, detectable modulation of the generator armature voltage resulted. For this reason, the philosophy of a "fast" regulator was abandoned, and it was decided that the frequency response of the system had to be lowered to assure better stability even with large noise inputs.

It was decided to introduce a "lag-lead"

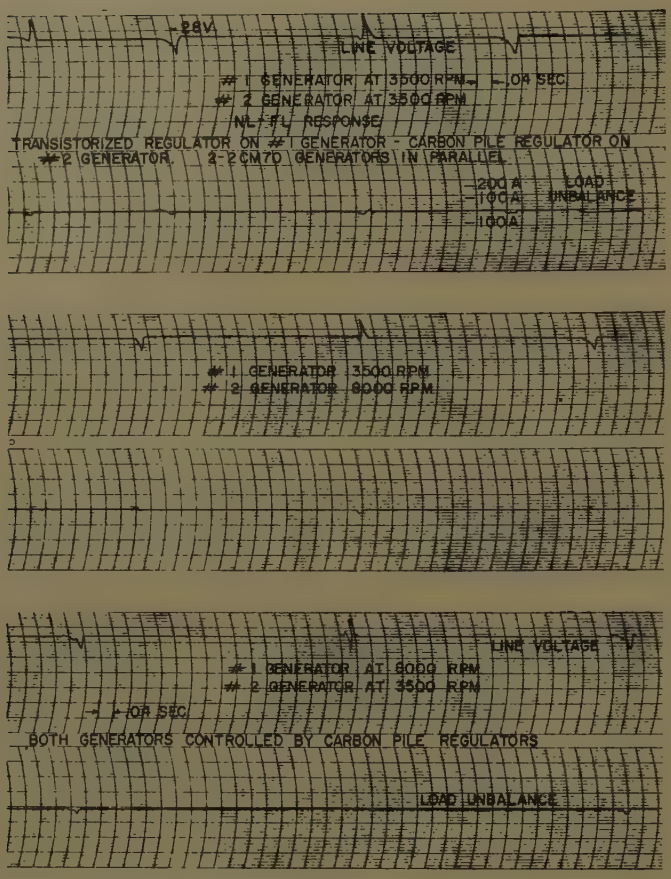


Fig. 13. Typical parallel operation of systems with one transistor regulator and one carbon-pile regulator demonstrating their compatibility (top). Carbon pile controlled systems (bottom) for comparison

characteristic to the existing regulator transfer characteristic. This was accomplished by the connection of the inductance $X1$ in series with the resistor $R15$ across a third control winding of the magnetic amplifier $SX1$. Fig. 9 shows the equivalent circuit of the magnetic amplifier which is obtained with this inductance stabilization scheme. The frequencies of the "break" points are:

$$\omega_1 = \frac{R_2}{L_2} \quad \omega_{21} = \frac{2}{b+\sqrt{b^2-4a}}, \quad \omega_{22} = \frac{b+\sqrt{b^2-4a}}{2a}$$

where:

$$a = \frac{L_1 L_2}{R_1 R_2}$$

$$b = \frac{L_1}{R_1} + \frac{L_1+L_2}{R_2}$$

Thus, it is possible to arrive at a precise design for the inductor, resistor, and associated magnetic amplifier winding, once it has been determined where to place the "break" points. Fig. 10 shows a Bode plot of the final regulator design. As can be seen from this figure, closed-loop response of the modified regulator has been reduced in frequency such that closed-loop gain is considerably less at the troublesome noise frequencies. This design was tested in the laboratory with outstanding success, generator field voltage remaining smooth for all conditions of speed and bus load.

Fig. 10 also shows that with generator field time constants varied over the extremes which might be expected from *MIL-G-6162* aircraft d-c generators, system stability is assured with an adequate phase margin in all cases.

The experience with this stabilizing problem has been very enlightening. It was found through this experience that a "slow" regulator must be used in this application in order to avoid the noise response difficulties which were experienced with the "fast" regulator circuit. It is very strongly felt that while a "fas regulator might be perfectly satisfacto in many applications, for a bona carbon-pile replacement regulator, sponse to the inherent and unpredictab high-frequency noise levels should be a tenuated.

6. Radio interference: A disadva tage of using a switching device for fie current control is that an appreciab amount of radio frequency noise is ge erated as a result of the steep current wa fronts involved during switching.

Radio noise testing was conducted a prototype regulator in accordance wit specification *MIL-I-6181B*, and it wa determined that filtering of the positiv power input lead and the field outpu lead was definitely necessary to enab meeting specification limits. Filters wer designed and installed (*C*1, *C*2, *C CH*1, *CH*2, referring to Fig. 6) and wer determined to be more than adequate t suppress radio noise below *MIL-I-6181* limits. *C*3 is not only a part of the radi noise filtering, but also provides protec tion for the regulator transistors since is large enough to filter voltage "spikes which may occur on the generator pos tive bus.

Electrical System Operation

Fig. 11 consists of ink oscillograp recordings showing typical single genera tor system performances of both the car bon-pile regulator and this propose transistorized replacement. Note partic ularly the excellent transient stabilit of the transistorized regulator as wel as the low initial overshoot durin system build-up.

Fig. 12 shows recordings in an actua system of parallel operation of transistor ized regulators. The excellent dynami transient stability should again be noted

The recordings of Fig. 13 indicate th compatibility of the transistorized regula tor with the carbon-pile regulator whe operating two generators in parallel. I is interesting to note that the transien stability of the system is improved ove that which was experienced with tw carbon-pile regulators controlling paral leled generators. Thus it is shown tha the addition of a transistorized regulato in a parallel generator system with othe carbon-pile regulators in this syste would result in an improvement of syste stability and performance.

Product Design

The major factors which dictated th progress of the product design were tha

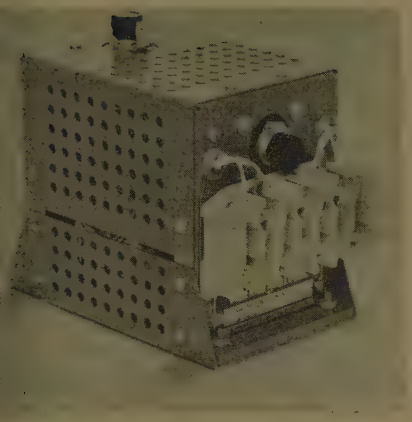


Fig. 14. Final product design

new regulator must be physically erchangeable with presently used car-n-pile regulators and must meet or ceed the same performance and en-onmental requirements, the most sig-icant of which is a 71 C ambient. ıxiliary requirements were maintain-ility and the closely associated char-teristics of high reliability and long life. ie need for improvement of these latter o properties over that of the carbon e is, of course, the major reason for this velopment and justification for its herently higher cost.

The matter of interchangeability is mplicated by the fact that the basic rbon-pile regulator is a plug-in unit ving two different mounting bases and tline requirements—one used by the nited States Air Force and the other by e United States Navy's Bureau of Aero-utics. A further complication is that e carbon pile is vibration isolated. A ggedized bolt-down design meeting the)-*g*-to-500-cycle requirement of the *IL-E-5272*, Procedure I, vibration ecification was considered necessary.

The unit shown in Fig. 14 has the quisite performance, including the ounting dimensions and outline, for the United States Air Force version. To accommodate the Bureau of Aeronautics (BuAer) version, the basic United States Air Force unit is provided with simple mounting feet extensions which attach to its original mounting holes. The equalizing control needed only in the BuAer version is added to the top, requiring only a slight modification to the upper cover. The advantages of having a basic design requiring only slight modifications such as this are obvious.

A terminal board, identical to that used on the BuAer base, is used in preference to a lighter weight, more convenient, and less expensive ON connector because of its greater reliability. Its location on the front accommodates present aircraft wiring while the location of the voltage-adjust control on the front provides good accessibility to it.

The use of germanium transistors, dictated by cost as well as other circuit considerations such as their low saturation resistance, resulted in one of the major product design problems, since the relatively low junction temperature tolerated by germanium necessitated their thermal isolation from the remainder of the circuitry. All five germanium transistors used in this design are located in the lower half of the package where both bottom and sides of the package are utilized for radiation. Heat transfer to them from the components in the upper half is effectively blocked from the lower half.

In addition, it was found necessary to thermally isolate several resistors, dissipating a total of approximately 20 watts, within a separate enclosure located on the rear of the package.

Tests in a large, thoroughly soaked chamber having no forced circulation indicate that the junction temperatures of the germanium transistors on this design are held to a safe level in an ambient of +71 C under the conditions requiring the highest dissipation within the regulator.

Construction details of interest are:

1. All electromagnetic components are cast with an epoxy resin in a single compact subassembly having self leads which act as a wire harness. This eliminates a large number of vulnerable connections and also provides a strong vibration-resistance building black.
2. Three filter capacitors are similarily cast in a subassembly.
3. All remaining components except the transistors and the voltage-adjust control are mounted on two parallel component boards located on either side of the electromagnetic and capacitor blocks. This provides for a strong assembly having a minimum of wiring and a maximum of accessibility.
4. Weight is less than 4 pounds.

Operational reliability, of course, cannot be determined until considerable field experience is accumulated; however, a paper estimate of expected reliability can be made. Using the exponential law of reliability and estimated failure rates for the various components, it is expected that the mean time between failures will be approximately 5,000 hours. This provides 99.9-per-cent reliability for a 5-hour mission and 90 per cent for 500 hours.

Conclusions

The development of this regulator clearly demonstrates that germanium power transistors can be applied in a properly designed regulator to enable direct replacement of 71 C carbon-pile regulators. Tests prove that this regulator exceeds the performance required by existing carbon-pile specifications.

Although initial cost of the transistorized regulator undoubtedly will be greater than the carbon pile, its longer life, increased reliability, and improved electrical performance should more than justify its application as a carbon-pile regulator replacement.

A New Brushless D-C Excited Rotating Field Synchronous Motor

GEORGE M. ROSENBERRY, JR.
MEMBER AIEE

SILICON RECTIFIERS have been used for several years to make brushless d-c excited rotating field synchronous generators.[1] The induced field current and voltage during starting have made similar brushless synchronous motors impractical. This paper describes the operation and performance of a new brushless synchronous motor which has been made practical by the development of high-current silicon-controlled rectifiers.[2] All sliding- and mechanical-type contacts have been eliminated. The synchronizing function and pull-out protection are automatically provided by a simple static circuit which uses two silicon-controlled rectifiers and Zener diodes. The new motor can be started like a squirrel-cage induction motor with a standard full or reduced voltage starter. The adjustable power factor correction of the conventional synchronous motor is retained.

This same development can be applied to brushless generators to provide complete protection to the rotating rectifiers for all conditions of operation including sustained out-of-step operation.

The Basic Motor Circuit

Principle of Operation

The basic synchronous motor circuit in its simplest form is shown in Fig. 1. The upper portion of the circuit, enclosed by a dotted line, includes all of the rotating components. The lower part shows the motor armature, exciter field, and rectifier d-c supply for the exciter field. Note that the exciter field is excited whenever voltage is applied to the motor.

During synchronous speed operation, excitation is supplied for the motor field as in a brushless generator. The exciter field is supplied with direct current by rectifying the motor line voltage. This field generates a voltage in the rotating a-c exciter armature. The voltage is then rectified by silicon rectifiers which are mounted on the rotor. The resulting direct current is fed directly to the motor field.

During the starting period, the motor field must be short-circuited to carry the induced field current. Otherwise the field voltage would be high enough to destroy the rectifiers. When terminal F_2 of the motor field is positive, the induced field current can flow through the exciter rectifiers, D_B and D_A. When terminal F_1 is positive, the current is carried through the controlled rectifiers SCR_A and SCR_B. Without these, there would be a large d-c component of current in the motor field, so that the motor would not start even though the field voltage was not sufficient to cause failur of any components. This is becaus the direct current in the field causes large braking torque when the speed o the rotor is low. The key to the operatio of the motor is in the control of the tw silicon-controlled rectifiers. They mus short-circuit the field during the startin period, but allow the application of direc voltage to the field near synchronou speed to pull the motor into step.

Detailed characteristics of silicon-con trolled rectifiers are given in reference They are basically a one-way switch Current flow is blocked in the revers direction as an ordinary rectifier. I the forward direction (direction of th arrow), they block current flow unt turned on by a small current injecte into the gate lead. When turned on they have the low forward voltage dro of a silicon rectifier. They remain i the conducting state until the curren through them is reduced to a sma value.

Zener diodes, Z_A and Z_B, are used a a voltage-sensitive device to turn on th controlled rectifiers at a fixed instantane ous positive field voltage. The volt ampere characteristic of a Zener diod is shown in Fig. 2. The current con ducted by the diode is very small unt the voltage exceeds the Zener voltag

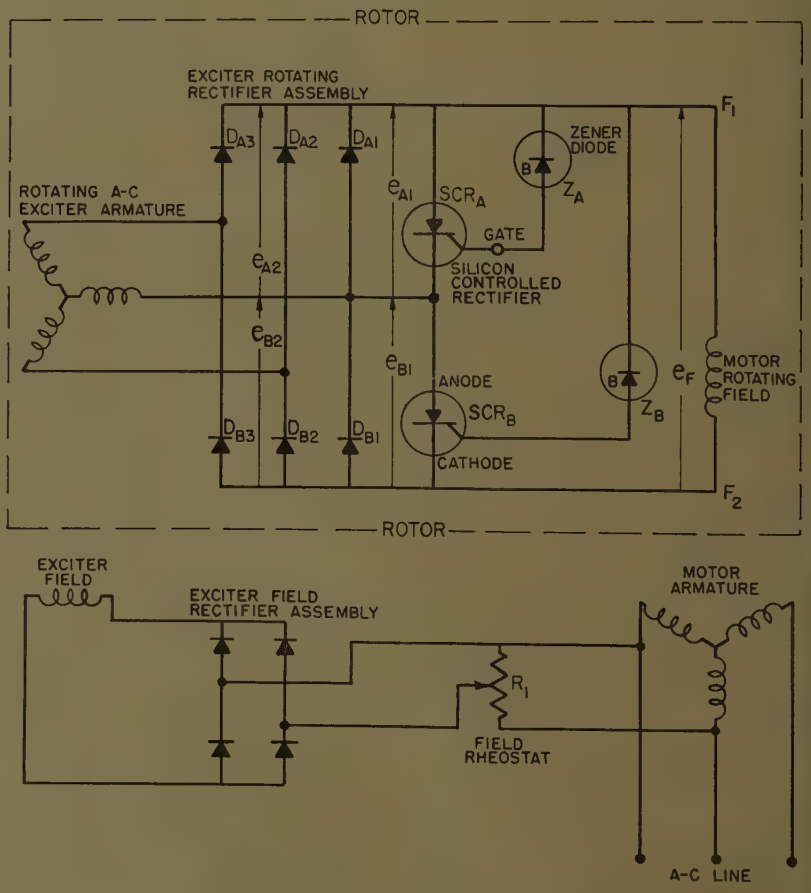
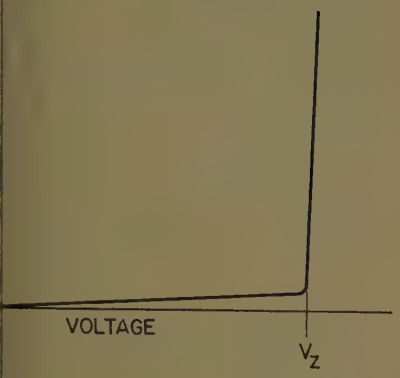


Fig. 1. Schematic diagram of brushless synchronous motor

Paper 60-127, recommended by the AIEE Rotating Machinery Committee of the Power Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 30, 1959; made available for printing November 27, 1959.

George M. Rosenberry, Jr., is with the General Electric Company, Schenectady, N. Y.

The author wishes to acknowledge that the basic idea to use two silicon-controlled rectifiers for short-circuiting the bridge-type rectifier was suggested by F. W. Gutzwiller.



2. Volt-ampere characteristic of Zener diode

ve that, a small increase in voltage lts in a large increase in current. 3 shows the silicon-controlled rectifier Zener diode used for tests described his paper.

he operation of the controlled-rectifier -shorting circuit is as follows:

s the voltage of motor field terminal, increases in the positive direction, current will flow through the Zener le, Z_B, until the Zener voltage is hed. Above that, a small increase voltage will result in enough gate ent to turn on the controlled rectifier. r SCR_B turns on, all of the field age will appear across SCR_A, and it turn on in a similar manner. This l turn-on time is about 10 micro-nds near zero speed with full voltage lied to the motor armature. This uit prevents the field voltage from eding the voltage at which the con-led rectifiers are turned on. The endix shows that the peak inverse age of the rectifiers can not exceed the k field voltage, so that complete pro-ion against excessive peak inverse age is provided.

ince the field is effectively short-uited, the motor will accelerate to speed. As it approaches full speed, exciter will start to deliver some ut current. This additional current carried by the controlled rectifiers l the motor starts to pull into syn-nism and reduces the induced field age. As the field voltage drops w the value required to trigger the trolled rectifiers, the exciter applies field to the motor, and it pulls into chronism.

Vith two controlled rectifiers con-ed across the field as shown in Fig. here is no possibility that the output ent of the exciter will keep them ed on. For example, when the ter is delivering direct current to the field circuit, diode D_{A1} conducts in the forward direction once each cycle. The forward drop of about 1 volt on the diode is applied in the inverse direction to controlled rectifier SCR_A. This turns it off. If only one controlled rectifier were used across the field, it could be held in the conducting state by the direct current output of the exciter.

Test Results

Tests were conducted on a 50-horsepower 1,200-rpm motor. The motor was coupled to a dynamometer with an inertia of 5.5 times the NEMA (National Electrical Manufacturers Association) normal value. Fig. 4(A) shows the induced field current waveform during the portion of the acceleration near synchronous speed. The motor was unloaded except for the inertia. It can be seen that the motor pulls into synchronism without any delay. Fig. 4(B) shows the field current with full-load torque applied during acceleration. Fig. 4(C) shows the field current as the motor is pulled out of step from full-load conditions. It can be seen that the field is immediately short-circuited on the first slip cycle. When the load torque is reduced, the motor will resynchronize.

Tests were conducted to determine the relative pull-in torque of the brushless synchronous motor compared to optimum angle application of the field with a separate source of field voltage.[3] The pull-in torque was 30% less than that obtained with optimum angle application. This value is adequate for many applications.

Discharge Resistor Shorting Circuit

In some cases, it is desirable to have an external resistor in series with the field during the starting period. This is needed to obtain improved accelerating torque. It is necessary to remove the discharge resistor from the circuit or short it, when the motor is synchronized, to prevent excessive losses. Fig. 5 shows how an additional silicon-controlled rectifier can be used to short-circuit the discharge resistor as the motor pulls into synchronism.

Fig. 3. Silicon controlled rectifier, rated 50 amperes (large device) and Zener diode rated 10 watts

Operation of Discharge Resistor Shorting Circuit

The two controlled rectifiers, SCR_A and SCR_B, short the output of the exciter, as was previously explained. Controlled rectifier SCR_D short-circuits the discharge resistor, R_D, after the motor comes up to speed and pulls into synchronism.

During the starting period when terminal F_1 of the motor field is positive, induced field current flows through the discharge resistor in the direction to impress inverse voltage across the controlled rectifier, SCR_D. It blocks current in this direction, so the discharge resistor carries all of the current. When terminal F_2 is

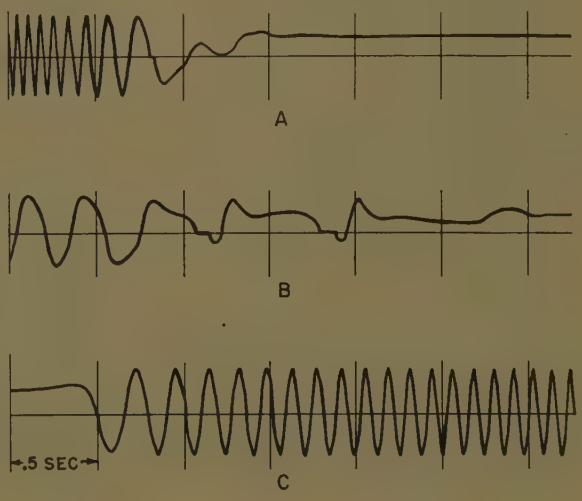


Fig. 4. Motor field current near synchronous speed for 50-horsepower 1,200-rpm motor

A—Acceleration with inertia load 5.5 times NEMA normal inertia

B—Acceleration with inertia load of 5 times motor inertia plus full-load torque

C—Pulling out of step from full load

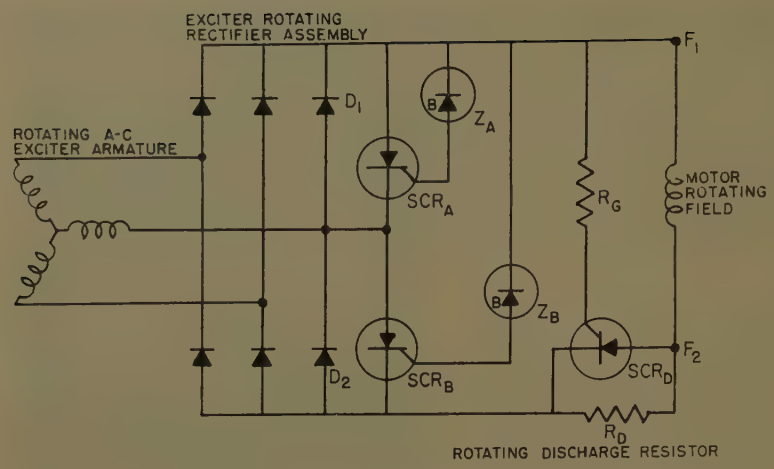


Fig. 5. Motor field circuit with discharge resistor shorting circuit

positive during the starting period, the current reverses and forward voltage is impressed on the controlled rectifier. It can not turn on, however, because the gate is about 2 volts negative with respect to the cathode. Therefore, during acceleration, the discharge resistor carries the induced field current 100% of the time.

When the motor comes up to speed and the exciter starts to deliver d-c output to the field, terminal F_1 of the field is positive. At the same time, forward voltage is impressed across the controlled rectifier. Since the gate supply voltage is positive, the controlled rectifier will turn on and remain on as long as direct current flows in the motor field.

Test Results

Tests on the 50-horsepower 6-pole motor confirmed the operation of the circuit as described. With a discharge resistance equal to twice the field resistance, this motor accelerated the 5.5 times normal inertia load in 70% of the time required without the discharge resistor.

Protection of Rotating Rectifiers on Brushless Synchronous Generators

The field shorting circuit discussed in this paper can be used to provide complete protection of generator rectifiers against overvoltage transients which may come from the generator field circuit under certain abnormal conditions of operation. The primary sources of these transients are:

1. Synchronizing with another generator out of phase or synchronizing in phase with the other generator voltage considerably higher.

2. Single phase line-to-line short circuit at the generator terminals, when the voltage is going through zero.

3. Sustained out-of-step operation with system.

4. If the generator is connected to a system and a short circuit close to the terminals is maintained long enough for the flux to decay, reapplication of system voltage in phase with the generator voltage will cause a negative field current.

Most small generators can be protected against all these conditions, except condition 3, with a small voltage-sensitive resistor. It is not generally necessary to protect against condition 3. Where it is necessary to protect against out-of-step operation for more than a few cycles, the controlled-rectifier field-shorting circuit offers a practical method, since the controlled rectifiers are small and dissipate very little heat. The difficulty with voltage sensitive resistors is that the voltage remains quite high, and this results in high power which can not be dissipated or stored if the generator is to be run out of step for more than a few cycles. Controlled rectifiers have a low impedance in the conducting state, so that the internal heat generation is relatively low.

Test results on a generator show that the field-shorting circuit provides adequate overvoltage protection to the rotating rectifiers. The field is shorted only during the time when high negative field current persists. The short is immediately removed, when it is no longer needed, to allow normal operation of the generator.

Conclusions

The brushless synchronous motor described in this paper does not require any auxiliary control for synchronizing, pull-out protection, or rectifier protection. These functions are all provided by a simple reliable circuit consisting of two silicon-controlled rectifiers and two Zener diodes. All sliding contacts, relays, contactors, and adjustments have been eliminated to provide utmost reliability and safety. The motor can be started like an induction motor with standard induction motor starters. The advantage of simple adjustment of power factor correction is retained.

This same field circuit can be applied to brushless generators to provide rectifier protection during sustained out-of-step operation

Appendix. Rectifier Peak Inverse Voltage Analysis

This appendix will show the limiting relationship between the field voltage and the rectifier peak inverse voltage.

The rotor circuit of a brushless synchronous machine is shown at the top of Fig. The sum of the voltages across the two rectifiers in any one leg of the bridge is the field voltage. Consider leg "2," for example:

$$e_{2A}+e_{B2}=e_F \qquad (1)$$

where

e_{A2} = voltage across diode D_{A2}. When e_{A2} is positive, inverse voltage is impressed on the diode

e_{B2} = voltage across diode D_{B2}

e_F = field voltage

If the forward voltage drop on the diode is taken as a maximum of 1.5 volts,

$$e_{A2}>-1.5 \qquad (2)$$

$$e_{B2}>-1.5 \qquad (3)$$

and

$$e_F>-3 \qquad (4)$$

This last equation shows that the field is essentially short-circuited in the reverse direction by the diodes.

If equation 3 is substituted into equation 1, the result is:

$$e_{A2}<e_F+1.5 \qquad (5)$$

Similarly:

$$e_{B2}<e_F+1.5 \qquad (6)$$

Positive voltage is inverse voltage on the diodes. These equations show that the maximum possible peak inverse voltage on the diodes is essentially equal to the peak positive field voltage.

References

1. An Oil-Cooled A-C Generator for High Speed High-Altitude Aircraft, H. J. Braun, W. J. Shillings. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 74, 1955 (Jan. 1956 section), pp. 456–60.

2. A Silicon-Controlled Rectifier—I. Characteristics and Ratings, D. K. Bisson, R. Dyer. *Ibid.*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 102–06.

3. Pull-In Characteristics of Synchronous Motors, D. R. Shoults, A. H. Lauder, S. B. Crary. *Ibid.* (*Electrical Engineering*), vol. 54, Dec. 1935, pp. 1385–95.

Discussion

. **Alger** (Consulting Engineer, Schenec-, N. Y.): This paper marks a very im-ant advance in the art and, in my ion, paves the way for a much greater of synchronous motors for industrial and r purposes.

the first place, this new brushless syn-nous motor can be built in totally en-ed fan-cooled and explosion-proof de-s, just as induction motors are. This not feasible heretofore, because of the blems of exposed sliding contacts and h wear.

the second place, the substitution of the dard and inexpensive full-voltage induc--motor starting panel, for the complex expensive synchronous motor starting el formerly required, should make the machine much more attractive commer-y than existing types of synchronous or. The automatic transfer of the field uit from d-c supply to short-circuit ugh a discharge resistor and the re-e, as may be required on starting and k overload conditions, eliminate the troublesome field relays and contactors which were required with synchronous motors.

In the third place, it should be feasible to design the new motors for lower starting currents than induction motors of comparable performance, since the synchronous machine amortisseur winding is not called upon to carry the running current. This should make the new machines attractive for air-conditioning applications, where many schemes, such as part-winding starting, have been employed to reduce the in-rush current.

Finally, because of the very small field excitation requirements, it should be feasible to apply automatic programming devices to this excitation supply. This should enable the motor to have high power factor for limited periods when charges due to a power factor clause make this desirable. It may also have high breakdown torque under short-time overload conditions. That is, the motor can be designed with a relatively small air gap and in an economical frame size, and yet have the output and high power factor of a considerably larger frame size for short periods, within the limits of its short-time heating capacity.

Of course, these expectations are contingent upon the satisfactory performance in the field and of the silicon-controlled rectifiers and Zener diodes, and on the assumption that these devices will not have an excessive cost.

G. M. Rosenberry, Jr.: I appreciate Mr. Alger's discussion, in which he has given an excellent appraisal of some of the potential applications of the new brushless motor.

Silicon-controlled rectifiers and Zener diodes have already been proved capable of long trouble-free life, when properly manufactured and applied. They do not have limited life as is the case with selenium rectifiers. Zener diode costs are already at a fairly low level. Silicon-controlled rectifier prices have been dropping rapidly. This is illustrated by the cost of the 16-ampere controlled rectifier, which has decreased in cost by a factor of 8 to 1 in the last 2 years.

Present silicon-controlled rectifiers are capable of handling a 1,200-h orsepower 6-pole synchronous motor, and larger controlled rectifiers are expected in the future.

ull-In Criterion for Reluctance Motors

JOHN F. H. DOUGLAS
FELLOW AIEE

HIS SUBJECT is timely because of the recent improvement of reluctance tors to a point where they have con-rable sales volume. The pull-in erion is related to the maximum rtia that can be pulled into step for a en frame size in the rating set on that ne, expressed as a graph or by an equa-. This paper derives this equation. reluctance motor is a synchronous tor without rotor excitation. The l-in criterion for synchronous motors h d-c excitation was studied by Edger- and Fourmarier,[1] and by Shoults, ry, and Lauder.[2] Although their lts cannot be applied to the reluctance tor directly, they were found useful in cking certain steps.

When the pull-in criterion is shown phically, each frame size will have ves as in Fig. 1. The abscissa denotes rating, in horsepower (hp), of the ne for high moment of inertia, J. us if a large inertia must be pulled into p, the rating must be reduced. The maximum horsepower shown is fixed by other properties of the design, such as heating, starting torque, or pull-out torque. The figure also shows how the slip varies with the rated load. The inertia increases as the rating is reduced, roughly as the inverse square of the rating. The curves show the upper limit for the inertia. The equations will be given in a later section.

Fig. 2 presents a graphic picture of the last instants before pull-in. It shows slip variation, and resultant torque variation, plotted against θ, the angle between the rotor interpolar axis, and its equilibrium position E. The slip is marked S and the resultant torque T_r. Curve *1* marks the condition of an inertia slightly greater than the critical value. Curve *2* is for an inertia slightly less than the critical value. Point E is a point of unstable equilibrium, so curve *2* shows a hunting oscillation before reaching the point of stable equilibrium E_2. Point F marks the position of the revolving field.

In the interval from the point of maximum slip and the equilibrium point E, the kinetic energy of the rotor must be increased. The area under the torque curve represents the apparent work done. These two must be related. Thus the method used in this study is based on the transformation of electric energy into kinetic energy. It should be noted here that the base B for a reluctance motor is 180 electrical degrees, and that the base N is one electrical radian. The area under the torque curve and the area of the rectangle are equal. The advantage of attacking the problem by energy relations is that it leads to an ordinary calculus solution. The solutions in references 1 and 2 used nonlinear equations, which was much more difficult.

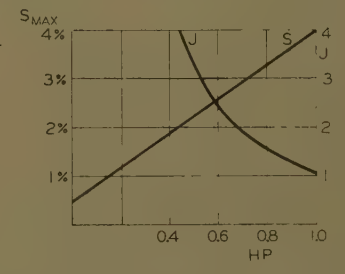


Fig. 1. Inertia and rating for a given frame

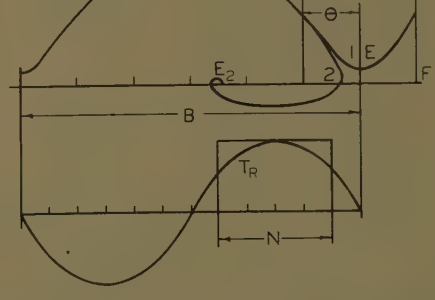


Fig. 2. Slip and torque curves near pull-in with an inertia near critical value. Curve 1 inertia is a little too large. Curve 2 is satisfactory and near limiting value

er 60-166, recommended by the AIEE Rotating chinery Committee of the Power Division and roved by the AIEE Technical Operations partment for presentation at the AIEE Winter eral Meeting, New York, N. Y., January 31–ruary 5, 1960. Manuscript submitted August 959; made available for printing December 10, 9.

N F. H. DOUGLAS is with Marquette University, waukee, Wis.

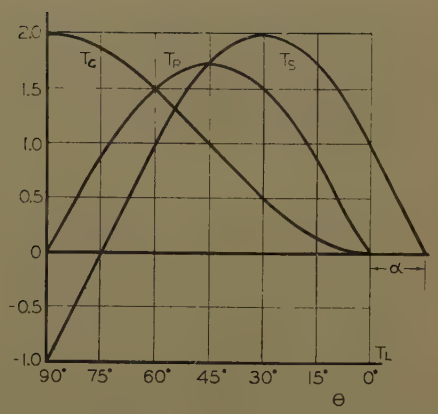


Fig. 3. Resultant and component torques. α is the angle between the synchronous and resultant torques

Although the motor parameters may be predicted by design methods, this paper assumes that the parameters of the circle diagram of the motor have been found. In checking the criterion of pull-in by test, a pen recorder is recommended, with a battery voltage in the pen circuit so that the pen records slip instead of speed (i.e., an oscillograph recording directly on paper tape). The maximum slip just before pull-in is to be noted. It may be stated that the criterion given here checks data furnished by the Louis Allis Company within the probable error of the test.

Components of Resultant Torque

During the last cycle of pull-in, the torque has three chief components: that produced by synchronous motor action, that produced by induction motor action, and the mechanical load and other constant retarding torques. These three components are shown in Fig. 3 as T_s, T_c, and T_L for the synchronous, cage, and mechanical torques. The figure is drawn for a typical motor at nominal rating and with critical inertia. It is to be noted that the maximum value of the resultant torque is approximately 87% of that due to synchronous motor action, and is the cosine of the angle α.

The minor corrections to the resultant torque obtained by a detailed consideration of the components raise the ratio of resultant torque to synchronous torque from the approximation shown in Fig. 3 to 0.88.

Succeeding sections of this paper state the four chief equations and give their proof, followed by test suggestions and several sections on the minor correction factors. One minor factor will be discussed further in the Appendix.

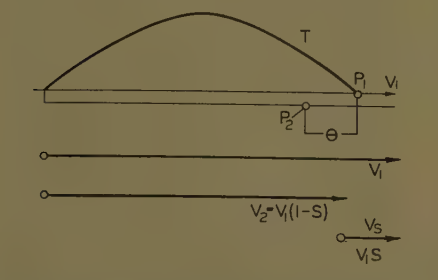


Fig. 4. Velocity diagram where P_1 and P_2 are corresponding points on the stator and rotor and θ is the slip angle

Nomenclature

E = line voltage
f = frequency
$f'\theta$ = function in equation 9
$f\theta$ = integral of $f'\theta$
G = conductance of cage
hp = rating of motor
I = motor current
I_1 = current defined by equation 8
J = moment of inertia
K_1 = ratio of resultant to synchronous torque
KE = kinetic energy
P_1 = input power
P_0 = core, friction, and windage
r = stator resistance, per unit (pu)
R = Radius of circle locus, pu
S = pu slip
T = pu torque
T_c = cage torque
T_L = equivalent load torque
T_m = mechanical load torque
T_r = resultant torque
T_s = synchronous torque
ΔT = minor torque correction
V_2 = velocity of rotor
V_s = slip velocity
W_a = apparent work
W_r = work on the rotor
X = reactance
ω_m = angular velocity in mechanical radians per second
α = angle defined by Fig. 3
δ = angle defined by Fig. 5
ϵ = angle OCA in Fig. 5
θ = angle between rotor and its equilibrium position

Statement of Formulas

Four chief formulas are to be proved: for kinetic energy, for apparent work, for the pull-in criterion, and for the shape of the wave of slip.

If the moment of inertia of the motor and load be expressed in kilogram-meters squared and called J, and if the angular velocity of the rotor at synchronous speed in mechanical radians per second be called ω_m, then the maximum kinetic energy of the system at synchronous speed in joules is given by equation 1:

$$KE_{\max} = \frac{1}{2}(J)(\omega_m)^2 \qquad (1)$$

The apparent work during the last half torque cycle, or the area under the torque curve in Fig. 2, when expressed in joules is given by equation 2:

$$W_a = (K_1)(E^2)(\text{hp})\left(\frac{746}{2\pi f}\right) \quad (R)$$

In equation 2 the constant K_1 is appr
mately 0.87, E is the pu voltage, hp is
nominal horsepower of the frame
is the frequency, and R is the pu valu
the radius of the circle locus.

The maximum moment of inertia wh
can be used at nominal horsepower
still pull into step, is given by the
equality 3 and depends on the maxim
slip and the quantities in equation
and 2.

$$(J)(\omega_m)^2(S_{\max})^2 \angle W_a$$

The shape of the slip curve in Fig.
given by equation 4 approximately

$$s = S_{\max} \sin(\theta)$$

Proof of Equations 1–4

The equation for kinetic energy
found in texts on dynamics written
the meter-kilogram-second system
units. In the conversion of electric
kinetic energy, the meter-kilogram-seco
system has the advantage of needing
conversion factor since both are in jou

Equation 2 with the omission of
factor $2\pi f$ is the torque in synchron
watts. Dividing by $2\pi f$ gives the tor
in newton-meters for a 2-pole mot
In Fig. 2 the base N is 1 radian, so t
the expression then gives work in jou
For a motor with p poles the torque
multiplied by $2p$ but the angle in
chanical radians is divided by the sa
factor. The poles, therefore, do
appear in the formula.

In Fig. 4 the point P_1 on the stato
traveling at synchronous speed. Po
P_2 on the rotor is moving at the speed
times that of the stator. The relat
speed or slip speed is s times the speed
the stator. The ratio of the rotor sp
V_r to slip speed V_s is given by equation

$$V_2/V_s = \left(\frac{1-s}{s}\right) = (W_r/W_a)$$

Since work is proportional to torque ti
velocity, the work on the rotor bears
same ratio to the apparent work as th
respective speeds. While the rotor
slipping one pole (as in the last half tor
cycle) the rotor may travel many po
It is important to note that the ene
W_a is converted into heat in the c
while the energy W_r is converted i
kinetic energy.

Now consider the changes in work a
kinetic energy during the interval wh
slip changes by ds and the angle chan
by $d\theta$:

$$E) = KE_{max}d(1-s)^2 = \left(\frac{1-s}{s}\right)(T_r)(d\theta)$$

$$= KE_{max}(2)(1-s)(-ds)$$

ltiplying both sides of the equation $s/1-s$ and taking the complete integral es equation 6:

$$_{max}(S_{max})^2 = W_a \quad (6)$$

en the values for kinetic energy and from equations 1 and 2 are used, the ılt is the pull-in criterion, equation 3. This procedure may also be used for the omplete integral. If the integration made in reverse order the left-hand mber of equation 6 has S instead of the ximum slip. The equation becomes

$$W_a \sin(2\theta)d\theta = W_a\left(\frac{1}{2}\right)(1-\cos 2\theta)$$

$$= W_a \sin^2\theta = (KE_{max})S^2 \quad (7)$$

e ratio of equation 7 to equation 6 gives shape of the slip wave given in equa- n 4, for inertias near the critical value. It is to be noted that while the minor ms of torque may modify equation 4 ghtly, they are of such a character that validity of equation 3 is not im- ired.

ıll-In Tests

In the discussion of the paper cited in ference 2 it was suggested that a battery included in the pen circuit of a pen corder, of such voltage that the slip ther than the speed is recorded. This important for accurate data. To check g. 1 of this paper a series of graded fly- heels should be used to supplement the ertia of the motor and dynamometer.

If the designer wishes to check his mputed cage resistance, two methods n be used. He may multiply the com- ted average slip by 1.57 which should eck the maximum measured slip before ll-in. This is based on the sinusoidal ape of the slip wave in Fig. 2. If a avy flywheel of about five times the itical inertia is used, the pen recorder ll show a wave of slip differing only out a ±10% from the average. Under ese conditions the space and time aver- es of speed are nearly equal, and the n recorder will show the average slip, be checked against computed cage sistance. Such studies may help to oduce a better cage.

Tests at 110% voltage should be made the motor is to be furnished with an totransformer. The motor is given an pulse of higher voltage just before ll-in. A study of equations 2 and 3 in- cates that this may increase the critical ertia by 45%.

When accurate tests have been made, the proposed pull-in criterion checked, and reliable design formulas for cage resistance found, then the realistic ratings of a given frame size under high inertia may be established.

Synchronous Torque

This and the next two sections will consider the minor factors affecting the constant K_1 in the pull-in criterion. This section is devoted to the synchronous torque. Since the last half-cycle of torque before pull-in takes about half a second, transients may be neglected. Thus the circle locus found after pull-in may be used before pull-in as well.

The circle locus of a reluctance motor, shown in Fig. 5, shows the phasor of current I in terms of the torque angle 2δ and also the input P_1. This may be established by the Blondel theory or by test. It is shown in sales literature and proved in a paper by the writer.[3] It will be assumed that the parameters I_0, R, and the stator resistance are known. The parameter I_1 may be computed by

$$I_1 = (r/X_dX_q) = I_0(I_0+2R)(r) \quad (8)$$

The gross torque is input minus stator copper loss. The loss is I^2r and the input is given by equation 9:

$$P_1 = I_1 + R \sin(2\delta)$$
$$= (I_0)(I_0+2R)(r) + R\sin(2\delta) \quad (9)$$

It is to be noted that in these and the following equations pu values are convenient. Unit voltage is rated voltage, unit current that of an ideal motor, and unit impedance is the ratio of these two.

The cosine law in trigonometry helps to find the copper loss. If the angle OCA be called ϵ, then the copper loss will be given by equation 10:

$$I^2r = (I_0+R)^2(r) + R^2r - 2(R)\times$$
$$(I_0+2R)(r)\cos(2\delta+\epsilon) \quad (10)$$

Both equations 9 and 10 have a constant term and a sinusoidal term. The constant term ΔT_L is largest and will be found first. It is found by combining equations 9 and 10:

$$\Delta T_L = -I_0(I_0+2R)(r) + (I_0+R)^2(r) +$$
$$R^2r = +2(R)^2(r) \quad (11)$$

For a typical motor with an R of 2 pu and an r of 0.025 pu, ΔT_L is (+0.20). Thus the effect of stator resistance is to add a constant load to the motor, so far as torque is concerned.

When the sinusoidal components of equations 9 and 10 are combined the net result is a rather small correction of approximately 1%. First evaluate the

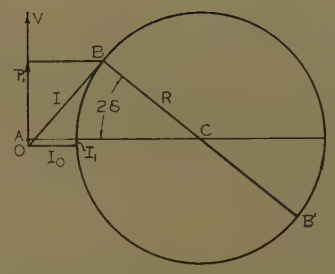


Fig. 5. Circle locus of a reluctance motor

very small angle ϵ which is approximately 2 degrees. Thus, assuming $\cos(\epsilon)$ is unity,

$$\tan(\epsilon) = I_1/(I_0+R) = (I_0r)\left(\frac{I_0+2R}{I_0+R}\right) \quad (12)$$

then

$$T_s = R\sin(2\delta) + 2R(I_0+R)(r)\cos(2\delta+\epsilon)$$
$$= R\sin(2\delta) + 2R(I_0+R)(r)\cos(2\delta) -$$
$$2R(I_0+R)(r)\sin(\epsilon)\sin(2\delta)$$
$$= R\sin(2\delta) - 2Rr(I_0+R)(I_0R)$$
$$\left(\frac{I_0+2R}{I_0+R}\right)\sin(2\delta) +$$
$$2Rr(I_0+R)\cos(2\delta)$$
$$= R\sin(2\delta)(1-2r^2I_0(I_0+$$
$$2R)) + 2Rr(I_0+R)\cos(2\delta) \quad (13)$$

The two terms in equation 13 are two perpendicular phasors and the cosine term is small, so that the hypothenuse may be found by the binomial theorem with the exponent (1/2), so that if the corrected value of R be termed R',

$$R'/R = 1 - 2I_0r^2(I_0+2R) + (2(I_0+R)r)^2$$
$$= 1 + 2r^2(I_0^2 + 2I_0R + R^2 - I_0^2 - 2I_0R)$$
$$= 1 + 2(Rr)^2 \simeq 1.005 \quad (14)$$

The sinusoidal wave is advanced in phase for a typical motor by 8 degrees. This does not affect the constant K_1.

Summarizing this section, the actual motor with resistance is equivalent to one without resistance but an added equivalent load. Thus the synchronous torque and the equivalent load are given in equations 15:

$$T_s = R'\sin(\Delta+8°) \quad (15A)$$

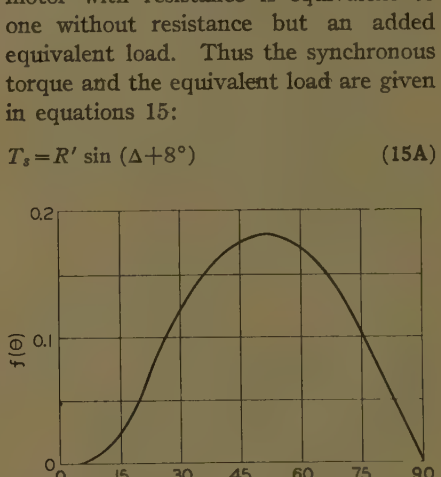


Fig. 6. Correction factor f(θ) for finding minor terms in slip and torque

$$T_L = T_m + \Delta T_L + P_0 \quad (15B)$$

Here T_m is the actual mechanical load, P_0 is core friction and windage, and ΔT_L is the equivalent added load. Equations 15 are in the pu system.

Cage Torque

In this section closer approximations are made to the torque produced by the slip, than were made in the first approximation shown in Fig. 3. One important principle is that the average cage torque must supply the load torque including friction and windage, and also the parasitic torque ΔT_L discussed in the last section. To obtain the second approximation for cage torque it will be assumed that torque and slip are proportional, and that the slip is governed by equation 4. Then it will be demonstrated that the cage torque contains another term varying with the sine of 2θ. This term is small but is needed to obtain an accurate value of K_1 appearing in the pull-in criterion.

One assumption is that the cage is isotropic so far as torque is concerned. This is proved by the small torque dip in these motors at half-speed. A brief analysis in the Appendix confirms this assumption. The ratio of cage torque to slip may be measured or it may be computed.

In Fig. 3 the various torques are related to the angle α and the quantity R'. The following equations apply:

$$T_L = (R') \sin(\alpha) \quad (16A)$$

$$T_r = (R') \cos(\alpha) \quad (16B)$$

The first approximation to the cage torque, shown in Fig. 3, is given by

$$T_c = 2T_L \sin^2\theta \quad (17)$$

The first approximation to the slip, given in equation 4, shows that the next approximation for cage torque must be given by

$$T_c = (\pi/2)(T_L) \sin(\theta) \quad (18)$$

It is to be noted that the factor $\pi/2$ assures an average value to balance T_L. The first correction to the resultant torque is the difference between the last two equations and may be denoted by ΔT_r. Thus

$$\Delta T_r = (T_L)(1.57 \sin(\theta) - 2.00 \sin^2\theta) = (T_L) f'(\theta) \quad (19)$$

Thus, when expressed in the pu system a closer approximation for the resultant torque is given by equation 20:

$$T_r = R' \cos(\alpha) \sin(2\theta) + \Delta T_r = R' \cos(\alpha) \sin(2\theta) + T_L f'(\theta) \quad (20)$$

This term $f'\theta$ will not affect the criterion given in equation 3 because its average value is zero. It does, however, affect the waveform of slip. Thus, with the reasoning used in deriving equation 7 and integrating T_r to obtain slip squared, equation 21 is obtained.

$$(S/S_{max})^2 = \sin^2(\theta) + (T_L \sec\theta/R') \int_0^\theta f'(\theta) d\theta \quad (21)$$

The integral of $f'\theta$ is shown in Fig. 6 and it is approximately equal to the value:

$$f(\theta) = 0.24 \sin\theta \sin 2\theta \pm 0.01 \quad (22)$$

Thus equation 21 may be written in the form given in equation 23:

$$(S/S_{max})^2 = \sin^2\theta + 0.24(T_L \sec(\alpha)/R') \sin\theta \sin 2\theta \quad (23)$$

Next, the square root is taken using the binomial theorem, giving equation 24:

$$(S/S_{max}) = \sin(\theta) + 0.12(T_L \sec(\alpha)/R') \sin(2\theta) \quad (24)$$

The wave of slip, when found experimentally, should show a second harmonic. The curves in reference 2 did show such a harmonic of small amount.

This component of slip makes an additional small contribution to the torque of the cage. Thus, if the right-hand member of equation 24 is substituted for $\sin\theta$ in equation 18 a revised value of cage torque is

$$T_c = (\pi/2)(T_L)(\sin(\theta) + 0.12(T_L \sec(\alpha)/R') \sin(2\theta) \quad (25)$$

The correction is important since it adds about 9% to the resultant torque.

The Constant K_1

This is the constant appearing in equation 2 and carried into equation 3. The earlier approximation appearing in Fig. 3 was 0.87. Since only terms in 2θ affect the complete integral, the corrected value of resultant torque combines the earlier value of $R'\cos(\alpha)$ with the second term of equation 25. This yields equation 26:

$$T_r = [R' \cos(\alpha) + 0.12(1.57)(T_L^2/R') \sec(\alpha)] \sin(2\theta) \quad (26)$$

Taking the value of R' as 1.005 times the radius of the locus R and dividing equation 26 by R' and omitting the factor $\sin 2\theta$, expression for K_1 is found and is given inequation 27:

$$K_1 = [\cos(\alpha) + 0.19(T_L/R')^2 \sec(\alpha)]\ 1.005 \quad (27)$$

Using typical motor constants and nominal frame rating, the constant is 0.88.

Equation 27 was used in computing the shape of Fig. 1. For example, if high inertia compels the rating to be set at 80% of nominal horsepower, the constant K_1 increases to 0.93.

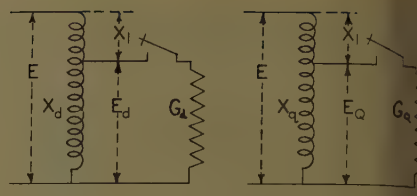


Fig. 7. Circuits for determining the lack isotropy in the rotor

Conclusions

1. A formula has been found for fin ing the maximum inertia that a relu tance motor will pull into step at any loa provided the designer has reliable formul for the cage resistance.

2. Tests are described for gettin accurate design constants, and for settin the rating on a given frame size for heav inertia duty.

3. Present cages are nearly isotrop so far as torque is concerned.

4. For the teacher, a discussion of pul in is given, which will supply a top missing in the textbooks. It is teachab since ordinary calculus is used. Whi developed for the reluctance motor, it easily extended to other synchrono motors.

Appendix

In reference 1 it was stated that if t rotor cage was not isotropic, the ca torque would vary according to the followin

$$T_c = K(1 - b\cos 2\theta)$$

In this equation the constant b marks t degree of anisotropy of the cage. T author consulted a paper by Linville,[4] an offers Fig. 7 as an aid for computing relati torques in the direct and quadrature pos tions. In Fig. 7, X' is the stator reactan and X_d and X_q are the direct and quadratu reactances. The conductance of the cage denoted by G which varies roughly with t amount of conducting metal on the tw axes. A ratio of 1.6 would make the rot isotropic as to torque. Data from co mercial motors indicate that they are near isotropic. An analysis of the factor indicates that its effect on K_1 is very small.

References

1. The Pulling into Step of a Salient-Po Synchronous Motor, Harold E. Edgerton, Pa Fourmarier. *AIEE Transactions*, vol. 50, Ju 1931, pp. 769–81.

2. Pull-In Characteristics of Synchrono Motors, D. R. Shoults, S. B. Crary, A. H. Laud *Ibid.*, vol. 54, Dec. 1935, pp. 1385–95.

3. Current Loci of Salient-Pole Synchrono Motors: An Extension of Blondel Theor J. F. H. Douglas. *Ibid.*, pt. III (*Power Apparat and Systems*), vol. 74, Dec. 1955, pp. 1080-83.

4. Starting Performance of Salient-Po Synchronous Motors, T. M. Linville. *Ibi* vol. 49, Apr. 1930, pp. 531–47.

Comparison of Vibratory and Rotating-Wheel Gyroscopic Rate Indicators

G. C. NEWTON, JR.
MEMBER AIEE

RECENT ADVANCES in atmospheric and space flight have increased the d for reliable and accurate instru- ntation for sensing the angular mo- ns of aircraft, missiles, and space ships. h measurements are especially needed guidance and control purposes. At present time, the preferred devices for asuring angular motion are based on rotating-wheel gyroscope. By the penditure of much research and de- opment effort these gyroscopic devices ve been brought to a state of perfection ich makes inertial navigation tech- ally feasible but very expensive. The sent reliance on the rotating wheel es not mean that other devices have not en considered. Over the years a num- of alternatives to the rotating wheel ve been devised and some of these have en developed to the point where prac- al limitations on performance were nd. In general, however, the total earch and development effort that s been devoted to these alternatives s been rather small compared with the ort that has been expended in the re- ement of the rotating-wheel gyroscope. Of the alternatives to the rotating- eel gyroscope those which are based the principle of vibratory rather than tary motion of mass elements represent important subclass. The tuning fork nfiguration of Fig. 1 is typical of this ss of instruments. In this paper we mpare the tuning fork device with the tating-wheel instrument of Fig. 2. order to simplify the discussion the mparison is made for instruments de- ned to measure rates of rotation with spect to inertial space. However, many the implications of this analysis also apply to instruments designed to measure angular displacements with respect to inertial space.

In forming a comparison between vibratory and rotating-wheel gyroscopic rate indicators it is assumed that the instruments are designed to convert angular rates into electric signals. Three quantities are considered in making the comparison. These are: 1. the sensitivity, in terms of output signal per-unit angular rate; 2. the thermal fluctuations (or noise) which are inherent in the measuring technique; and 3. the ratio of the thermal fluctuations to the sensitivity. This ratio is equivalent to an angular-rate noise which is superposed on the angular rate to be measured by the instrument. The last two quantities express the ultimate performance limitations which are inherent in these instruments. At present, certain practical problems occur in both classes of instrument; these practical difficulties cause relatively large errors which mask the ultimate performance limits. The fact that both classes of instrument have errors so large compared with the ultimate performance limits indicates that there remains much room for further improvement in instrument performance. Since considerable research effort has been devoted to the improvement of the rotating-wheel gyroscopes, it is recommended that a much more intensive program than has ever existed in the past be devoted to the overcoming of the practical difficulties in vibratory-type instruments. It is the purpose of this paper to call attention to the potential improvements in vibratory instruments that are possible providing the practical problems encountered in their realization can be overcome. Even though there is a similar margin of potential improvement for the rotating-wheel gyroscope the practical problems, particularly that of mass unbalance, are of such a nature that they may forever prevent the attainment of the performance levels indicated by the thermal fluctuation limits. In spite of the tremendous research efforts devoted to the rotating-wheel devices, the rate of improvement has been declining ever since the jump that occurred with the introduction of the flotation principle[1] for reducing friction levels. Apparently, a point of diminishing returns has been reached for the rotating wheel; therefore, in the allocation of future research efforts more consideration should be given to vibratory and other unconventional gyroscopic instruments.

Sensitivity

The tuning-fork form of vibratory gyroscopic rate indicator has been described in the literature[2-4] and therefore we shall give only a cursory description of its operation. Referring to Fig. 1 it is assumed that means which are not shown have been provided to keep the tines of the fork in oscillation. The vibrating tines result in a time-varying moment of inertia about the z-axis. If the instrument as a whole is rotating around the z-axis with an angular rate Ω with respect to inertial space, conservation of angular momentum requires that a vibratory torque exists on the base supporting the tines. Mounting the base on a spring suspension which has a natural frequency equal to that of the tines permits the vibratory torque to set up an angular oscillation of the base around the z-axis that will be of measurable magnitude. The oscillatory motion of the base is assumed to be detected by means of a variable mutual inductance as shown in Fig. 1. The primary of the mutual inductance is supplied with a direct current I_1. Thus the output signal e_2 of the secondary will be an alternating voltage

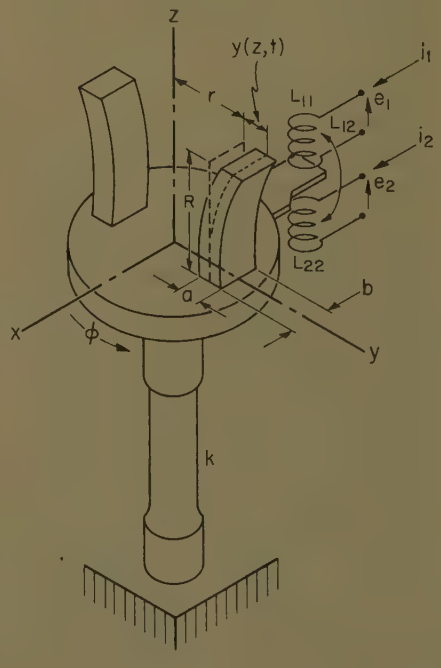


Fig. 1. Tuning fork configuration of vibratory rate gyroscope

per 60-62, recommended by the AIEE Feedback ntrol Systems and approved by the AIEE Tech- al Operations Department for presentation at AIEE Winter General Meeting, New York, Y., January 31–February 5, 1960. Manuscript mitted September 24, 1959; made available for nting November 25, 1959.

C. NEWTON, JR., is with the Massachusetts stitute of Technology, Cambridge, Mass.

e research for this paper has been supported in rt by the Lake States Oil Company, First Na- nal Bank Building, Iron Mountain, Michigan, a rporation for whom the author consults. The thor also expresses his appreciation to his col- gues at the Massachusetts Institute of Tech- logy who have patiently discussed with him ne of the ideas expressed herein.

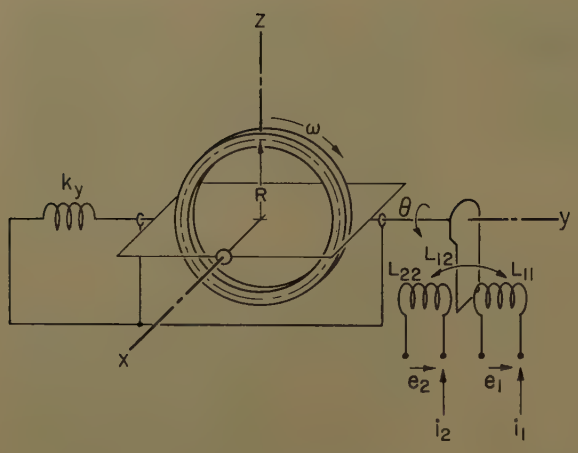


Fig. 2. Rotating-wheel rate gyroscope

of frequency equal to the base vibration frequency.

A sensitivity figure for the tuning-fork embodiment of the vibratory class of rate indicators is now defined. Simply taking the open-circuit output voltage from the secondary of the pickup and dividing this by the angular rate being measured will not be satisfactory as a figure for sensitivity.

It is obvious that the impedance of the electric circuit into which the output voltage will be fed must be taken into account. When this is done the source impedance of the instrument itself becomes important. The source impedance is made up of several components. First, there are those associated with the pickup such as the secondary resistance and self-inductance. Second, there is the impedance of the mechanical system as reflected into the secondary circuit. Here only the latter component will be considered and the sensitivity S_v for the vibratory instrument shall be defined to be

$$S_v \triangleq \frac{(E_2/\Omega)}{\sqrt{R_{m\phi}}} \tag{1}$$

where E_2 is the amplitude of the open-circuit pickup voltage, Ω is the angular rate of the instrument as a whole around the z-axis, and $R_{m\phi}$ is the resistance of the mechanical system associated with the sensing axis as reflected into the pickup secondary circuit. For maximum power transfer into the electric circuit, the resistance facing the pickup should be equal to $R_{m\phi}$ if it is assumed that the winding resistance of the pickup secondary is negligible and that the effects of inductances have been removed by tuning. This definition of sensitivity is independent of transformation ratio on either the mechanical or electrical sides of the pickup and is therefore believed to be superior to that employed by Chatterton.[3]

On the basis of the foregoing definition an expression is sought for the sensitivity in terms of the basic dimensions and parameters of the instrument. Assume the tines of the fork to be rectangular prismatic bars with the dimensions shown in Fig. 1. The tines of the fork are assumed to have a mass m; this mass is considered to be the active mass of the instrument. The amplitude of vibration of the tines is limited by the maximum strain ξ_{mv} that can be permitted in the material used to construct them. In order to account for the speed of response of the instrument, a time constant τ is defined as the rate of change of the phase of the modulation envelope of the pickup output signal with respect to frequency for sinusoidal oscillation of the instrument about the z-axis. An attempt to maximize the sensitivity of the instrument by adjusting its dimensions under the constraint of fixed time constant and fixed active mass m shows that the sensitivity is substantially independent of all instrument dimensions (and therefore of its operating frequency). As shown in the Appendix, the sensitivity turns out to be

$$S_v = \frac{0.441\tau^{1/2}m^{1/2}\left(\frac{E}{\rho}\right)^{1/2}\xi_{mv}}{\alpha_\phi{}^{1/2}} \tag{2}$$

Here E is Young's modulus and ρ is the density of the active material. $(E/\rho)^{1/2}$ is the velocity of sound for extensional waves. The quantity α_ϕ is the ratio of the total moment of inertia around the sensing axis to that represented by the tines alone when they are in their neutral positions.

This inertia ratio accounts for the inertia of the pickup moving element and of the base to which the tines are fastened. Normally it would be expected to have a value in the range of three to five.[3] The sensitivity has the dimensions of power$^{1/2}$ multiplied by time; in the meter-kilogram-second system of units this turns out to be watts$^{1/2}$ times seconds. The numerical constant in the sensitivity expression is dimensionless.

Now turn to the derivation of a corresponding sensitivity figure for the conventional rotating-wheel rate gyroscope. A rotating-wheel rate gyroscope is shown in Fig. 2. This rate gyroscope operates on the principle that torque is required to change the direction of angular momentum of a rotating wheel. Thus if the instrument as a whole has an angular rate with respect to inertial space around the z-axis, a torque will have to be exerted on the gimbal which supports the wheel. By mounting the gimbal in bearings and using a relatively stiff spring constraint, the required torque is developed by deflection of the gimbal from its neutral position. This deflection in the steady state becomes a measure of angular rate. In order to avoid an excessively oscillatory transient response to a suddenly imposed angular rate, it is customary to provide viscous damping in addition to the spring restraint. Some form of pickup means is usually provided in order to convert the deflection of the gimbal into an electric signal. For purposes of this discussion it will be assumed this pickup means to be a mutual inductance of the same type as is used in the illustration of the vibratory rate-measuring instrument. The theory of the rotating-wheel rate gyroscope is extensively treated in the literature,[1,5] so that certain well-known results will be used without derivation.

The sensitivity of the rotating-wheel instrument shall be defined in a way which is quite analogous to the definition used for the vibratory instrument. In the case of the rotating-wheel instrument, however, it is necessary to supply the primary of the mutual inductance used for the pickup with a-c excitation for otherwise there would be no secondary output signal under steady-state conditions. Assume positive directions for the instantaneous values of current and voltage to be as shown in Fig. 2. Furthermore, assume that a displacement of the gimbal in the positive direction θ as shown in this figure will yield a secondary voltage e_2 which is leading the primary current i_1 by 90 degrees when the secondary is under no load. If a current i_2 flows in the secondary circuit in the direction indicated and this current is in phase with the open-circuit voltage e_2, analysis indicates that there is no average torque reaction on the pickup armature. Thus there can be no resistive component to the impedance reflected into the electric circuit from the mechanical system. However, if the secondary current i_2

the open-circuit secondary voltage
by 90 degrees, thereby placing it in
se with the primary current i_1, there
be an average torque acting about
gimbal axis in such a direction as to
ease the angle θ and thereby increase
Thus the impedance reflected from
mechanical system into the electric
uit appears to be inductive. The
uctive reactance $X_{m\theta}$ is defined to
the reactance of the reflected induct-
e that corresponds to the pickup
itation frequency. Superficially, it
uld appear that one could draw in-
initely large amounts of power from
pickup providing one kept the
ondary current i_2 exactly in phase
180 degrees out of phase) with the
tage e_2. However, any small de-
rture in phase angle would immediately
ult in substantial torque reactions on
gimbal and depending upon whether
phase deviation is in the leading or
ging direction there could be either
unstabilizing influence or an attenuat-
effect on the gimbal displacement.
erefore, it is sound engineering prac-
e to have the impedance level of the
plifying equipment connected to the
kup secondary bear a definite relation
the inductive reactance reflected from
e mechanical system. Thus, from the
int of view of instrument sensitivity
e is justified in regarding the inductive
actance $X_{m\theta}$ for the rotating-wheel
strument as substantially analogous
the resistance $R_{m\theta}$ which was reflected
om the mechanical side of the vibratory
strument. Hence, the sensitivity S_r
the rotating-wheel rate gyroscope
defined to be

$$\frac{\Delta(E_2/\Omega)}{\sqrt{X_{m\theta}}} \qquad (3)$$

here E_2 is the amplitude of the open-
cuit pickup voltage and Ω is the
gular rate of the instrument as a
ole around the z-axis.

Using the definition of sensitivity pre-
ously given there is derived in the
pendix an expression which relates
e sensitivity to performance and ma-
ial parameters, and to a characteristic
nension of the rotating-wheel rate
roscope. In deriving this expression
tain assumptions are made. It is
sumed that the frequency of the
ernating current impressed upon the
mary of the pickup is the same as the
n frequency of the rotating wheel.
oser agreement with current practice
uld be obtained if we merely assumed
e pickup excitation frequency to be
portional to the spin frequency; this
uld merely introduce a frequency
proportionality factor into the sensitivity expression. The wheel is assumed to be a massive rim which is thin compared with its radius R. It is assumed that there is negligible mass in the web of the wheel. Thus the active mass m is the mass of the rim. The speed of response of the instrument is accounted for by a time constant τ which is the rate of change of phase of the gimbal tilt angle θ with respect to frequency for sinusoidal variation of the angular rate about the z-axis. τ is the negative slope of the phase versus frequency characteristic in the vicinity of zero frequency. It is also assumed that the maximum tilt angles of the gimbal which are encountered are small. Using these assumptions it is possible to arrive at the following expression for sensitivity

$$S_r = \frac{0.707\tau m^{1/2}\left(\frac{E}{\rho}\right)^{3/4}\xi_{mr}{}^{3/4}}{\zeta\alpha_y{}^{1/2}R^{1/2}} \qquad (4)$$

Here α_y is the ratio of the total moment of inertia associated with the gimbal axis to the moment of inertia of the wheel. This inertia ratio will customarily range from three to five. ζ is the damping ratio which characterizes the transient response of the gimbal deflection θ. This damping ratio normally ranges from 1 down to 3/10. The square root of E/ρ represents the velocity of sound in the rim material of the wheel. ξ_{mr} represents the maximum strain that is permitted in the rim material. The dimensions of the sensitivity S_r are the same as before and the numerical constant carries no dimensions.

Compare the sensitivities of the rotating-wheel and vibratory rate-measuring instruments. If it is assumed that the inertia ratios are the same for both the vibratory and rotating-wheel instrument and that the damping ratio of the rotating-wheel rate gyro is approximately unity then the sensitivity of the rotating-wheel instrument relative to the vibratory instrument is given by

$$\frac{S_r}{S_v} = \frac{1.60\tau^{1/2}\left(\frac{E}{\rho}\right)^{1/4}\xi_{mr}{}^{3/4}}{R^{1/2}\xi_{mv}} \qquad (4A)$$

Here ξ_{mr} is the maximum strain allowed in the rim material of the rotating wheel and ξ_{mv} is a maximum strain permitted in the tines of the vibratory instrument. In arriving at this sensitivity ratio it is assumed that the rim and the tine materials are identical so that the velocity of sound in each is the same. However, the same assumption cannot be made for the maximum strains for reasons which are discussed in the "Example Section" which follows. To give an idea of the numerical value of this ratio for typical values of the parameters, assume τ to be 0.06 second; R, 0.02 meter; $(E/\rho)^{1/2}$, 5,000 meters per second (this corresponds to steel); ξ_{mr}, 2.3×10^{-5}; ξ_{mv}, 1×10^{-3}. For these values the sensitivity ratio turns out to be 65.2 in favor of the rotating-wheel rate gyroscope. On the surface, this would seem to indicate a serious handicap for the vibratory instrument. However, before drawing this conclusion one must consider other performance factors than sensitivity. Although a high sensitivity is desirable since it reduces the amount of amplification required in the equipment which uses the output signal, sensitivity alone is not a sufficient criterion for judging the value of an instrument.

Ultimate Performance Limitation

In practical instruments there are many factors which tend to degrade performance. In the conventional rotating-wheel gyroscopes zero stability is one of these. Zero shifts can be caused by mass unbalance in the presence of gravitational or acceleration fields. Such mass unbalance with respect to the gimbal axis can result from lubricant shift in the ball bearings, wear, deflection of parts under combined acceleration and gravitational fields, dimensional instability of the materials of construction, and thermal distortion effects, to mention only a few of the more obvious. However, suppose that by great skill in instrument construction the zero shift problem is overcome. There would still remain the problem of thermal fluctuations; thus as a performance limitation thermal fluctuations appear to be quite fundamental. Therefore thermal fluctuations will be considered to be the ultimate performance limitation on gyroscopic rate-measuring instruments.

It is well known from general thermodynamical considerations that any system which is in thermal equilibrium with its environment will exhibit fluctuations of the state variables which determine the energy of the system.[6] In particular, if the component of the system's energy that is associated with a state variable is a quadratic function of that variable, then the variable must have a mean-square value such that the average value of the energy component is equal to $KT/2$ where K is Boltzmann's constant and T is the absolute temperature of the system and its surroundings. In the case of the vibratory rate-measuring instrument the dynamical system to be

considered is that of the sensing axis. The rms value of the thermal fluctuation voltage induced in the pickup under open-circuit conditions is designated by the symbol σ_{nv}. To be consistent with our definition of sensitivity given by equation 1 this rms noise signal must be normalized by dividing by the square root of the resistance $R_{m\phi}$ reflected into the electric circuit from the mechanical system. Thus we arrive at a normalized fluctuation voltage or noise N_v given by the definition

$$N_v \triangleq \frac{\sigma_{nv}}{\sqrt{R_{m\phi}}} \tag{5}$$

This expression can be evaluated in terms of the instrument parameters by equating the average kinetic energy of the sensing axis to $1/2\ KT$. As shown in the appendix the result turns out to be

$$N_v = \sqrt{\frac{2KT}{\tau}} \tag{6}$$

This result is rather surprising since it involves only one parameter of the instrument: its time constant τ which is inversely related to its speed of response.

From the viewpoint of what we are ultimately trying to measure, namely, angular rates, the critically important measure of performance is the instrument's noise level in terms of an equivalent rms angular rate $\sigma_{\Omega v}$ that would be necessary at the input of a noise-free instrument in order to produce the same output fluctuation as exists in the real instrument. This quantity is obtained by dividing the noise by the sensitivity or in terms of symbols

$$\sigma_{\Omega v} \triangleq \frac{N_v}{S_v} \tag{7}$$

It should be noted that this figure is completely independent of the pickup means and is therefore a very fundamental performance limitation. In practical instruments there will be other sources of noise besides the mechanical elements of the measuring system. For example, one of these additional sources is the pickup and associated amplifying means. However, by appropriate design it should be possible to limit the total noise to an rms value of the order of $1^1/_2$ to 3 times as large as that given by equations 6 or 7.

The noise equivalent rms rate is evaluated by dividing the right member of equation 6 by the right member of equation 2 thereby obtaining

$$\sigma_{\Omega v} = \frac{3.20(\alpha_\phi KT)^{1/2}}{\tau m^{1/2}\left(\frac{E}{\rho}\right)^{1/2}\xi_{mv}} \tag{8}$$

This result is believed to be an expression of the most fundamental performance limitation inherent in the tuning-fork configuration of vibratory rate-measuring instruments.

The rms noise in the output signal of the rotating-wheel rate gyroscope is defined in a manner similar to that used for the vibratory instrument. Specifically

$$N_r \triangleq \frac{\sigma_{nr}}{\sqrt{X_{m\theta}}} \tag{9}$$

where σ_{nr} is the rms modulation of the suppressed-carrier open-circuit output signal from the pickup. This noise quantity is proportional to the rms value of the gimbal deflection θ. This in turn is related to the mean energy stored in the gimbal restraining spring which is equal to $KT/2$. Following the steps shown in the appendix leads to

$$N_r = \frac{(2KT)^{1/2}\left(\frac{E}{\rho}\right)^{1/4}\xi_{mr}{}^{1/4}}{R^{1/2}} \tag{10}$$

as an expression for the normalized thermal fluctuation noise in terms of the basic instrument parameters. This result turns out to be somewhat more complex than the corresponding expression for the vibratory instrument. It involves the radius R of the rim of the rotating-wheel as well as two material parameters: the velocity of sound $(E/\rho)^{1/2}$ and the maximum permitted value of strain ξ_{mr}. But in contrast to the vibratory instrument it does not involve the speed of response.

Again consider the basic performance limitation to be the noise equivalent rms rate $\sigma_{\Omega r}$ which is obtained by taking the ratio of the noise N_r to the sensitivity S_r; this definition is analogous to equation 7. For the rotating-wheel rate gyroscope division of the right member of equation 10 by the right member of equation 4 yields for the noise equivalent rms rate

$$\sigma_{\Omega r} = \frac{2\zeta(\alpha_y KT)^{1/2}}{\tau m^{1/2}\left(\frac{E}{\rho}\right)^{1/2}\xi_{mr}{}^{1/2}} \tag{11}$$

It is interesting to note that all instrument parameters enter into the expressions for the noise equivalent rates given by equations 8 and 11 in exactly the same way with the single exception of the maximum permissible strain. There is also a difference in the numerical constants.

Examples

In order further to compare the vibratory and rotating-wheel instruments let us compute values of the equivale rms neois rates. In each case this w be done for reasonable values of t parameters. Consider first the vibrato instrument. Let the following values the parameters be employed: $\alpha_\phi =$ $K = 1.38\times10^{-23}$ joules per degree Kelv $T = 300$ degrees Kelvin, $\tau = 0.06$ second, $= 0.026$ kilogram, $(E/\rho)^{1/2} = 5{,}000$ mete per second, $\xi_{mv} = 0.001$. Substituti of these values of the parameters in equation 8 yields

$$\sigma_{\Omega v} = \frac{3.20[4(1.38\times10^{-23})300]^{1/2}}{0.06(0.026)^{1/2}5000(0.001)}$$

$$= 0.851\times10^{-6}\ \frac{\text{radians}}{\text{second}} \tag{1}$$

The value of maximum strain ξ_{mr} is base on a limited amount of experience wi vibratory instruments together wi fatigue estimates for alloy steels. Sin it is normal to design vibratory instr ments for frequencies in the low aud range the allowable stress limits ha to be set for such a large number cycles that for all practical purposes w are working with indefinitely long lives

In the case of the rotating-whe instrument the maximum permitte strains in most designs are well belo the limit which have been specified f the vibratory instrument. The reas for using such low strains appears be associated with ball-bearing li problems and unbalance problems ass ciated with nonisotropic expansion und the influence of centrifugal forces. Ther fore, the maximum value of strain sha be derived by reference to design e perience with conventional gyroscope It is easily shown that the strain in th rim of our idealized wheel is given by

$$\xi_{mr} = \frac{\omega^2 R^2}{\left(\frac{E}{\rho}\right)} \tag{1}$$

Here ω is the angular velocity of rotatio in radians per unit time and R is th radius of the rim. A typical rotor one with the radius of 0.02 meter an an angular velocity of the order of 1,20 radians per second. The strain in th rim material for these conditions is

$$\xi_{mr} = \frac{[1200(0.02)]^2}{(5000)^2} = 2.3\times10^{-5} \tag{1}$$

if it is assumed that the rim material steel in which the velocity of sound 5,000 meters per second. Thus it seen that conventional gyroscopes a designed for relatively low levels maximum strain.

The remaining parameters needed calculation of the noise equivalent rm

e are assigned values as follows: α_y 4, $K=1.38\times10^{-23}$ joules per degree lvin, $T=300$ degrees Kelvin, $\tau=0.06$ ond, $m=0.026$ kilogram, $(E/\rho)^{1/2}=$ 00 meters per second, $\zeta=1.0$. With se values substituted into equation one finds the noise equivalent rms e to be

$$\dot{\gamma}=\frac{2(1)[4(1.38\times10^{-23})300]^{1/2}}{0.06(0.026)^{1/2}5000(2.3\times10^{-5})^{1/2}}$$

$$=1.11\times10^{-9}\frac{\text{radians}}{\text{second}} \quad (15)$$

is obvious that the thermal fluctuation nit indicated by the result of equation for the rotating-wheel gyroscope is not immediate concern to designers of day's gyroscopes since zero-shift and her difficulties are normally several ders of magnitude above this level. he threshold of the model *JRT* inneapolis Honeywell Rate Measuring yroscope is advertised in the manucturer's bulletin *JRT 6-57* to be 0.01 egrees per second which is equivalent 1.75×10^{-4} radians per second.) Even bring the drift rate associated with ass unbalance to within a factor of ten the noise equivalent rms rate would quire the center of gravity offset of mbal system to be no more than $.36\times10^{-10}$ meters assuming the comined gravity and acceleration fields re one g. (This calculation assumes the mbal mass is α_y times the wheel mass; his should be a reasonable approximaion.) This offset is comparable to the iameter of electron orbit of the hydroen atom. Also in the case of the ibratory instrument there are correponding zero-shift problems associated ith the cross-coupling difficulties beween the vibrations of the tines and the scillations about the sensing axis.[4] hus designers of these instruments lso find that the ultimate limit on perormance represented by the noise equivaent rms rate is far from an immediate oncern. However, if some way is evised for considerably reducing or liminating zero-shift difficulties, then he ultimate performance limitations ndicated by our examples in equations 2 and 15 may become very significant. t should be noted that the rotatingheel device enjoys less than an order f magnitude advantage over the vibraory instrument with respect to thermal luctuations. When one considers that he zero-shift problem may be easier to vercome in the case of the vibratory nstrument than in the case of the otating-wheel device, the desirability f increasing the research and developent effort devoted to the vibratory forms of instrument becomes quite evident.

For the benefit of the reader who may doubt the relevance of a thermal fluctuation limit on performance for an instrument such as a gyroscope, the experience of galvanometer designers should be recalled. Ingenious ways for amplifying the response of galvanometers have been devised and used. These developments have been carried to the point where thermal fluctuations of the basic mechanical movement became the limiting factor.[7] If this limit was reached in the case of the galvanometer, who is to say that it may not some day be reached in the case of the gyroscope? After all, both the vibratory and rotating-wheel rate measuring instruments contain mechanical systems which are in some respect similar to that of the galvanometer.

Conclusion

This paper has derived sensitivity expressions for both the vibratory and rotating-wheel gyroscopic rate indicators. In addition, it has also postulated thermal fluctuations to be the ultimate limitation on performance and has developed expressions for the thermal noise limits inherent in both classes of instrument. Although the rotating-wheel device enjoys a relatively large advantage over the vibratory instrument from the viewpoint of sensitivity, it has much less of an advantage with respect to thermal noise limitations. Furthermore, the thermal noise limitation on performance for the vibratory instrument is considerably lower than the center-of-mass fluctuation and other limits encountered in rotating-wheel gyroscopic devices. If, perchance, it should prove to be possible to overcome the cross-coupling problems encountered in the vibratory instrument and thereby to overcome the equivalent zero-shift difficulty in this instrument, then the performance of such instruments may be improved to the point where thermal noise becomes a practical consideration. If this should come to pass the vibratory instrument could very well displace rotating-wheel devices in many applications. In the opinion of the author, it is unlikely that the zero-shift problem of the conventional gyroscope, which is primarily associated with mass unbalance, will ever be reduced to levels such that the thermal fluctuation limit becomes an important factor. This is based both on the physics of the situation as well as on an extrapolation of zero-shift improvement over the years. Since vibratory instruments have not been as fully explored as conventional gyroscopes, it is recommended that a greater research effort be devoted to them. This research should have the objective of reducing or eliminating cross-coupling problems in order to raise the performance until it approaches the thermal fluctuation limit.

Appendix

Derivation of Sensitivity of Tuning Fork Instrument

The driving torque M about the sensing axis is given by

$$M=\frac{d}{dt}(J_a\Omega) \quad (16)$$

where J_a is the moment of inertia of the active material about the z-axis and Ω is angular rate of the instrument as a whole which is assumed to be about the z-axis. It will be assumed that the tines of the fork are vibrating sinusoidally so that the instantaneous deflection of any point on the tine (see Fig. 1) is

$$y(z,t)=y(z)\sin\omega t \quad (17)$$

Then, assuming that the width b of the tines is small compared to the distance r of the neutral plane of the undeflected tine from the z-axis

$$J_a=2\int_0^R dz\rho ab[r+y(z,t)]^2 \quad (18)$$

where ρ is the density of the tine material, a is the thickness of tine, and R is its length. (The factor of two accounts for the second tine.) Thus, ignoring the effects of variations in Ω which will be slow compared to the vibration frequency of the tines, for the driving torque one has

$$M=2\Omega\frac{d}{dt}\int_0^R dz\rho ab[r+y(z)\sin\omega t]^2 \quad (19)$$

Or dropping second order terms in $y(z)$

$$M=4\Omega\omega\rho abr\cos\omega t\int_0^R dzy(z) \quad (20)$$

The driving torque M acts on the tuned mechanical system which senses the pulsating moment. The sensing system, composed of the total inertia J_ϕ of the tines and the base, together with the torsion bar of stiffness k, is tuned to have a natural frequency nearly equal to the tine frequency ω. Damping d is also provided in order to control the speed-of-response. If one ignores the slight variation in J_ϕ caused by the fork vibration one can write as the equation of motion of the sensing system

$$\left[J_\phi\frac{d}{dt}+d+\frac{k}{\left(\frac{d}{dt}\right)}\right]\dot{\phi}=M \quad (21)$$

where $\dot{\phi}$ is the angular velocity of the fork base about the z-axis which is superposed on the motion of the instrument as a whole. This can be put in the form

$$\left[\frac{\left(\frac{d}{dt}\right)+\frac{1}{Q_\phi}+\frac{\omega}{\left(\frac{d}{dt}\right)}}{\omega}\right]J_\phi\omega\dot{\phi}=M \tag{22}$$

Here it is assumed that the natural frequency of the sensing system $\sqrt{k/J_\phi}$ is exactly equal to the fork frequency ω, i.e.

$$\omega=\sqrt{\frac{k}{J_\phi}} \tag{23}$$

Also Q_ϕ is defined as

$$Q_\phi\triangleq\frac{\sqrt{J_\phi k}}{d} \tag{24}$$

For a constant angular rate Ω the driving torque is a pure cosine wave of frequency ω. Because the driving frequency is equal to the natural frequency, the response ϕ will be a cosine wave in phase with the driving torque. By equations 20 and 22 for the amplitude $|\phi|$ one has

$$|\phi|=\left[\frac{4Q_\phi\rho abr\int_0^R dzy(z)}{J_\phi}\right]\Omega \tag{25}$$

But J_ϕ can be expressed in terms of the mean value of the active inertia $\bar{J}_a$ by means of the ratio α_ϕ.

$$\alpha_\phi\triangleq\frac{J_\phi}{\bar{J}_a} \tag{26}$$

Since $\bar{J}_a$ is simply $2\rho abR\,r^2$ equation 25 can be written as

$$|\phi|=\left[\frac{2Q_\phi\int_0^R dzy(z)}{\alpha_\phi rR}\right]\Omega \tag{27}$$

Now $y(z)$ must be determined. The differential equation of the cantilever tine is

$$EI\frac{\partial^4 y}{\partial z^4}=-\rho ab\frac{\partial^2 y}{\partial t^2} \tag{28}$$

with boundary conditions

$$\left.\begin{aligned} y=0,\ \frac{dy}{dz}=0 \quad &\text{at } z=0\\ \frac{d^2y}{dz^2}=0,\ \frac{d^3y}{dz^3}=0 \quad &\text{at } z=R\end{aligned}\right\} \tag{29}$$

In order to avoid the complexities of the exact solution we shall use the solution form of equation 17 with a Rayleigh approximate solution for $y(z)$. Specifically take

$$y(z)=y_R\left(1-\cos\frac{\pi z}{2R}\right) \tag{30}$$

With this approximate solution

$$\int_0^R dzy(z)=\left(1-\frac{2}{\pi}\right)Ry_R \tag{31}$$

so that equation 27 becomes

$$|\phi|=\left[\frac{2\left(1-\frac{2}{\pi}\right)Q_\phi y_R}{\alpha_\phi r}\right]\Omega \tag{32}$$

Now find y_R in terms of the maximum allowable strain ξ_{mv}. The maximum strain occurs at the root of the tine. From beam theory as applied to the rectangular tine one has

$$\xi_{mv}=\frac{aM_m}{2EI} \tag{33}$$

where M_m is the maximum bending moment and I is the moment of inertia of the cross section. But

$$M_m=EI\left(\frac{d^2y(z)}{dz^2}\right)_{\max} \tag{34}$$

so that for the assumed form of the deflection curve (equation 30)

$$\xi_{mv}=\frac{\pi^2}{8}\frac{ay_R}{R^2} \tag{35}$$

The pickup shall be assumed to be a mutual inductance with a d-c excitation I_1. Then the output voltage amplitude E_2 will be

$$E_2=\left(\frac{\partial L_{12}}{\partial\phi}I_1\right)|\dot{\phi}| \tag{36}$$

If a pickup current i_2 exists it will cause a torque on the sensing axis M_{i2} given by

$$M_{i2}=\frac{\partial L_{12}}{\partial\phi}I_1 i_2 \tag{37}$$

If the current is sinusoidal and has the frequency ω this torque produces an in-phase motion

$$\phi=\frac{Q_\phi}{J_\phi\omega}M_{i2} \tag{38}$$

This motion in turn produces a voltage e_2 given by an equation similar to equation 36. Thus the impedance met by the current i_2 on account of the mechanical system associated with the pickup is a resistance $R_{m\phi}$ given by

$$R_{m\phi}=\frac{Q_\phi}{J_\phi\omega}\left(\frac{\partial L_{12}}{\partial\phi}I_1\right)^2 \tag{39}$$

From the definition of sensitivity S_v given by equation 1 and using equation 36 one finds

$$S_v=\frac{J_\phi^{1/2}\omega^{1/2}}{Q_\phi^{1/2}}\left[\frac{|\phi|}{\Omega}\right] \tag{40}$$

From equation 26

$$J_\phi=2\alpha_\phi\rho abRr^2 \tag{41}$$

Combining equations 32, 35, 40, and 41 yields

$$S_v=\left[\frac{16\left(1-\frac{2}{\pi}\right)\sqrt{2}}{\pi^2}\right]\times\left[\frac{Q_\phi^{1/2}R^{5/2}(\rho b\omega)^{1/2}\xi_{mv}}{\alpha_\phi^{1/2}a^{1/2}}\right] \tag{42}$$

The Q_ϕ of the mechanical system is related to the speed-of-response of the instrument to changes in angular rate Ω. Ultimately the output signal e_2 is demodulated and the time constant τ of the instrument shall be defined to be equal to the negative rate of change of phase with respect to modulation frequency of the demodulated output signal for sinusoidal variation of Ω. Modulation theory tells us that, with the sensing axis natural frequency set equal to the fork frequency ω, this rate of change of phase is equivalent to the slope of the phase of the sensing axis response to sinusoidal torques at frequencies near ω. From equation 22 the slope of the phase of ϕ for variations $\Delta\omega$ of the frequency from ω is easily found and this is equated to $-\tau$. Thus

$$\tau=\frac{2Q_\phi}{\omega} \tag{43}$$

Finally, the frequency ω must be found in terms of the tine dimensions. By Rayleigh's method the natural frequency of the tines is found by equating the maximum kinetic energy T_m to the maximum potential energy V_m. From

$$dT_m=\frac{1}{2}\rho ab(dz)\omega^2y^2(z) \tag{44}$$

one has

$$T_m=\frac{1}{2}\rho ab\omega^2\int_0^R dzy^2(z) \tag{45}$$

and from beam theory

$$dV_m=\frac{EI}{2}\left(\frac{d^2y(z)}{dz^2}\right)^2dz \tag{46}$$

which yields (with $I=a^3b/12$)

$$V_m=E\frac{a^3b}{24}\int_0^R dz\left(\frac{d^2y(z)}{dz^2}\right)^2 \tag{47}$$

Carrying out the integrations for the assumed deflection curve of equation 30 and equating the two energies yields

$$\omega=\frac{\pi^2}{\sqrt{\left(\frac{3}{4}-\frac{2}{\pi}\right)768}}\sqrt{\frac{E}{\rho}}\frac{a}{R^2} \tag{48}$$

Recall that the active mass m is defined as

$$m=2\rho abR \tag{49}$$

Combining equations 42, 43, 48, and 49 yields the expression for sensitivity given by equation 2.

Sensitivity of Rotating-Wheel Rate Gyroscope

The spring-restrained rate gyroscope of Fig. 2 has a torque M acting on the gimbal axis of

$$M=J_r\omega\Omega \tag{50}$$

where J_r is the rotor inertia, ω is spin angular velocity, and Ω is the angular rate along the z-axis of the instrument. It is assumed that the gimbal tilt angle θ is small so that $\cos\theta\cong1$ and $\sin\theta\cong0$. If it is assumed that the rotor is a thin rim of radius R and mass m

$$J_r=mR^2 \tag{51}$$

This mass m corresponds to the active mass of the tuning fork.

The differential equation governing the gimbal tilt is

$$\left(J_y\frac{d^2}{dt^2}+d_y\frac{d}{dt}+k_y\right)\theta=M \tag{52}$$

where J_y is the inertia about the gimbal axis, d_y is the damping coefficient, and k_y

e spring rate of the suspension. This
e written as

$$\left(\frac{d^2}{dt^2}+\frac{2\zeta}{\omega_n}\frac{d}{dt}+1\right)\theta=\frac{M}{k_y} \tag{53}$$

e the damping ratio

$$\frac{d_y}{\sqrt{J_y k_y}} \tag{54}$$

the natural frequency

$$\sqrt{\frac{k_y}{J_y}} \tag{55}$$

steady-state tilt angle θ is given by

$$\frac{I}{} \tag{56}$$

ning the time constant τ characteriz-
the speed of response in analogous way
hat used for the vibratory device (viz.,
he negative rate of change of the phase
at low frequencies for sinusoidal torques)
has

$$\frac{2\zeta}{\omega_n} \tag{57}$$

J_y is expressed in terms of J_r by means
n inertia ratio α_y where

$$\frac{J_y}{J_r} \tag{58}$$

n one may write the following expression
the ratio of the steady-state tilt angle θ
he angular rate Ω being measured

$$\frac{\tau^2\omega}{4\zeta^2\alpha_y} \tag{59}$$

re equations 50, 55, 56, 57, and 58 have
n used.
Assuming that a mutual inductance is
d to convert angle to output-signal as
wn in Fig. 2, it is necessary that a-c
itation be supplied. Let the excitation
rent be sinusoidal with amplitude I_1 and
quency equal to the spin frequency ω.
en the output voltage amplitude E_2 is

$$=\frac{\partial L_{12}}{\partial\theta} I_1\omega \tag{60}$$

ere L_{12} is the mutual inductance which is
umed to vary linearly with the angle θ.
e output voltage will lead the excitation
rent by 90 degrees in the absence of load-
. Next it is noted that a sinusoidal load
rent of amplitude I_2 in the pickup sec-
lary which is in phase with the excitation
rent will cause an average torque on
gimbal equal to

$$=\frac{\partial L_{12}}{\partial\theta}\frac{I_1 I_2}{2} \tag{61}$$

r the directions shown in Fig. 2 this torque
l increase the angle θ by M_1/k_y and there-
produce a voltage amplitude change of

$$_2=\frac{\left(\frac{\partial L_{12}}{\partial\theta}\right)^2 I_1^2\omega I_2}{2k_y} \tag{62}$$

us the steady-state impedance looking
o the output terminals of the pickup is
luctive and equal to

$$X_{m\theta}=\frac{\left(\frac{\partial L_{12}}{\partial\theta} I_1\right)^2\omega}{2k_y} \tag{63}$$

This impedance is that portion of the total output impedance which is attributable to the mechanical system.

Using the sensitivity definition of equation 3 and combining equations 59, 60, and 63 to find that

$$S_r=\frac{\tau^2\omega^{3/2}k_y^{1/2}}{2\sqrt{2}\,\zeta^2\alpha_y} \tag{64}$$

But $k_y^{1/2}$ is equal to

$$k_y^{1/2}=\frac{2\zeta}{\tau}\,\alpha_y^{1/2}m^{1/2}R$$

on the basis of equations 51, 55, 57, 58. Thus

$$S_r=\frac{\tau\omega^{3/2}m^{1/2}R}{\sqrt{2}\,\zeta\alpha_y^{1/2}} \tag{65}$$

Finally, ω can be expressed in terms of the maximum allowable strain ξ_{mr} in the active material by equation 13. This yields

$$\omega=\frac{\left(\frac{E}{\rho}\right)^{1/2}\xi_{mr}^{1/2}}{R} \tag{66}$$

Placing this value of ω into equation 55 yields the expression for sensitivity S_r given by equation 4.

Thermal Fluctuations in Gyroscopic Rate Indicators

As discussed in the main body of this paper, thermal fluctuations are postulated to be the ultimate performance limitation of gyroscopic rate-measuring instruments. According to Boltzmann's statistics a system whose energy is a quadratic function of its state variables has a mean thermal energy component of $KT/2$ associated with each state variable if it is in thermal equilibrium with its environment. Here K is Boltzmann's constant and T is the absolute temperature. The sensing axis of the tuning fork instrument stores energy in two forms: kinetic and potential. Only the kinetic energy is of interest and this is given by $(J\dot{\phi}^2)/2$. Equating the mean value of this energy to $KT/2$ yields

$$\overline{\dot{\phi}^2}=\frac{KT}{J_\phi} \tag{67}$$

where $\overline{\dot{\phi}^2}$ is the mean square value of $\dot{\phi}$. The rms thermal fluctuation voltage induced in the pickup has been defined to be σ_{nv}; by equations 67 and 36 one finds

$$\sigma_{nv}=\sqrt{\frac{KT}{J_\phi}}\left(\frac{\partial L_{12}}{\partial\phi} I_1\right) \tag{68}$$

The resistive impedance reflected into the electric system is given by equation 39. Eliminating Q_ϕ/ω by equation 43 yields

$$R_{m\phi}=\frac{\tau}{2J_\phi}\left(\frac{\partial L_{12}}{\partial\phi} I_1\right)^2 \tag{69}$$

Division of the rms thermal fluctuation voltage σ_{nv} by the square root of this impedance gives the normalized noise signal N_v of equation 6.

A similar analysis of thermal fluctuations is now made for the rotating-wheel rate gyroscope. Again the sensing (gimbal) axis stores energy in kinetic and potential forms but in this case it is the potential energy $(k_y\,\theta^2)/2$ that is of interest. Equating the mean value of this energy to $KT/2$ yields

$$\overline{\theta^2}=\frac{KT}{k_y} \tag{70}$$

In this instance we have defined σ_{nr} to be the rms value associated with thermal fluctuations of the amplitude of the suppressed carrier signal derived from the pickup. From equations 70 and 60

$$\sigma_{nr}=\sqrt{\frac{KT}{k_y}}\left(\frac{\partial L_{12}}{\partial\theta} I_1\omega\right) \tag{71}$$

The inductive reactance reflected into the electrical system is given by equation 63. Division of the rms thermal fluctuation voltage amplitude σ_{nr} by the square root of this reactance yields for the normalized noise signal

$$N_r=\sqrt{2KT\omega} \tag{72}$$

Substitution of the value of ω given by equation 66 into this expression yields equation 10.

References

1. Instrument Engineering (book), **C. S. Draper, Walter McKay, Sidney Lees.** McGraw-Hill Book Company, Inc., New York, N. Y., vol. I, 1952; vol. II, 1953; vol. III, 1955.

2. New Space Rate Sensing Instrument, **J. Lyman.** *Aeronautical Engineering Review*, New York, N. Y., vol. 12, Nov. 1953, pp. 24–30.

3. Some General Comparisons Between the Vibratory and Conventional Rate Gyroscope, **J. B. Chatterton.** *Journal of Aeronautical Sciences*, New York, N. Y., vol. 22, 1955, pp. 633–38.

4. Zero Signals in Sperry Tuning Fork Gyrotron, **C. T. Morrow.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 27, 1955, pp. 581–85.

5. The Floating Integrating Gyro and its Application to Geometrical Stabilization Problems on Moving Bases, **C. S. Draper, W. Wrigley, L. R. Grohe.** *SMF Fund Paper no. FF-13*, Institute of the Aeronautical Sciences, Inc., New York, N. Y., 1955.

6. Atomic Physics (book), **G. P. Harnwell, W. E. Stephens.** McGraw-Hill Book Company, Inc., 1955.

7. A Natural Limit for the Sensibility of Galvanometers, **G. Ising.** *Philosophical Magazine*, London, England, vol. I, 1926, p. 827.

Discussion

J. B. Chatterton (Moeller Instrument Company, Inc., Richmond Hill, N. Y.): The author's interest and appreciation of the vibratory rate gyro is very encouraging, as it has been my opinion for some time that this type of gyro has not received the attention it deserves, both from theoretical as well as experimental considerations. This is still the adamant opinion of my former colleagues J. Lyman and R. Barnaby who pioneered the development. It is hoped this paper will help stimulate further work.

I would like to point out that the author's sensitivity factor would be especially useful for design of a high-efficiency vibratory gyro where output loading would provide the damping required for some suitable time constant. This, of course, would not be true for the rotating gyro.

The author's attempt to predict the ultimate accuracy of these devices is very interesting to me since I have recently taken another approach. A general theory was developed[1] which qualitatively compared measuring methods when all the system components were at a near-idealized state. The most accurate system proved to be the one with the least number of balances or reference changes of physical quantities. When applied to gyros,[2] the tuning-fork gyro was shown to be the optimum of known gyro types but not necessarily the ultimate in theory.

REFERENCES

1. THE UNCERTAINTY OF MEASUREMENT SYSTEMS, J. B. Chatterton. *Transactions*, Professional Group on Instrumentation, Institute of Radio Engineers, New York, N. Y., vol. PGI-7, Mar. 1958, pp. 90–94.

2. GYRO CONFIGURATION REQUIREMENTS FOR OPTIMUM ACCURACY, J. B. Chatterton. *Proceedings*, Sixth Annual East Coast Conference on Aeronautical and Navigational Electronics, Institute of Radio Engineers, Oct. 1959, pp. 7.61–7.66.

G. C. Newton, Jr.: The author wishes to thank Mr. J. B. Chatterton for bringing to his attention another viewpoint on accuracy limitations in angular motion sensing instruments. As expressed in Mr. Chatterton's references this viewpoint appears to be as follows. Any measuring system can be regarded as a cascade of elementary components. Each component has the possibility of introducing "reference changes" from sources within the component itself or through environmental influences. These reference changes, through their fluctuations, produce instability in the zero of the output indication. An example of a reference change for a rate gyroscope is mass balance in the presence of a gravitatio field. Mr. Chatterton concludes that preciseness of instrument indication will improved as the number of refere changes that are capable of influencing output is reduced.

The author agrees with Mr. Chatter that an instrument designer should atte to minimize the number of reference chan acting upon the output. However, it is important to consider the quantitative fects of each reference change since so may produce little, and others very la influence on the output indication. Usu it is reference changes or disturbances t act near the input end of an instrument t cause the great difficulty. This paper a lyzes one of these, the influence of ther noise, since this would exist even if others were eliminated. But this does mean that other reference changes are important in many practical instrume as Mr. Chatterton points out in his re ences.

Characteristics of Centrifugal Pumps and Compressors Which Affect the Motor Driver Under Transient Conditions

H. A. WIEGAND
NONMEMBER AIEE

LEE B. EDDY
MEMBER AIEE

THE BEHAVIOR of electric motor drivers under transient conditions is presenting an increasing number of problems to electrical engineers in the refining industry. The problems are compounded as larger machines are installed and as continuity of service through momentary voltage failures becomes more important to process operation.

The motor designer and the refinery electrical application engineer are most frequently concerned with the determination of starting torque, accelerating time, motor heating, relay settings, surge currents, system voltage drop, and other pertinent factors associated with the starting and reaccelerating of electric motors. However, before an accurate analysis can be made of motor and electric system behavior during various types of transient conditions, it is often essential that the moment of inertia (WK^2) and speed-torque characteristics of the driven machine be established. The WK^2 of a particular machine can usually be learned from the manufacturer, but speed-torque characteristics for various types of fluid and mechanical systems often require a more arduous investigation. Although the problems of starting large motors have been emphasized in recent literature, the intimately related subjects of pump and compressor characteristics which affect the motor driver have usually been ignored or oversimplified.

This paper presents in general terms the typical WK^2 values and the speed-torque characteristics of centrifugal pumps and compressors which occur during normal starting periods and immediately following momentary voltage failures to the driving motors. In the interest of simplicity, the paper utilizes a number of approximations, but the resulting data are sufficiently accurate for most transient studies. A substantial amount of background information is included, from which detailed analyses for specific pr lems can be derived. Throughout paper the reader is assumed to possess elementary knowledge of fluid mecha and centrifugal machines.

Paper 60-16, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted February 13, 1959; made available for printing October 7, 1959.

H. A. WIEGAND is with the Ingersoll Rand Company, New York, N. Y., and LEE B. EDDY is a consulting engineer, Mount Prospect, Ill.

Pumps and Liquid Systems

LIQUID SYSTEM CHARACTERISTICS

To establish the torque requirem of a pump at any particular instant, is necessary to know certain charact istics of the liquid system in which it operating. Fundamentally, these ch acteristics take the form of different pressure across the pump as a function liquid flow through the pump.

For a system in which a pump transf liquid from one elevation to a higher e vation, and in which pipe friction a velocity head are negligible, the requi pump differential pressure is essentia constant at all rates of flow, as illustra by curve *A* of Fig. 1. The difference elevation is termed "static head" a may be converted to differential press in pounds per square inch by calculat the weight of a column of the liquid inch square having the same height as difference in elevation. In some ca an effective static head exists where pump works in a system with a control

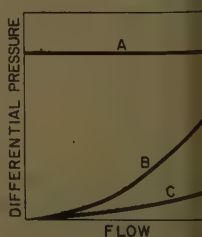
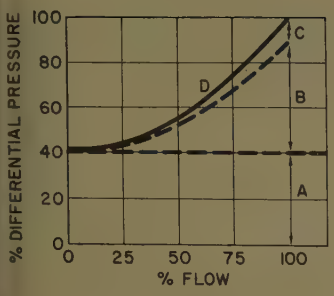


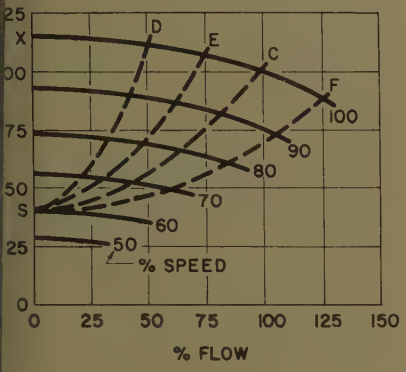
Fig. 1. System pressure - flow characteristics

A—Static head
B—Friction loss
C—Velocity head



2. Composite liquid system characteristics

A—Static head
B—Friction loss
C—Velocity head
D—Composite characteristics



3. Combined pressure-flow characteris- for a radial-flow pump and variable liquid em with 40% static head. System flow at following per cents:

C—100%
D—50%
E—75%
F—125%

k pressure, such as a system with a eiver maintained at constant pressure. ump feeding water to a boiler steam m is a familiar example of a system h back pressure controlled at a constant ue.

n a system where liquid is transferred m one location to another at the same vation and in which velocity head is gligible, the difference in pressure be- en the two locations is solely a result friction from the liquid acting against pipe, fittings, valves, and other restric- ns. In such a system, the required np differential pressure varies approxi- tely as the square of the flow, as shown curve *B* in Fig. 1.

n a frictionless system with no differ- e in static head, there is a component pump differential pressure termed locity head" which is proportional the square of flow, as illustrated by ve *C* in Fig. 1. In most practical sys- ns, velocity head is relatively small npared with the pressure loss resulting m friction.

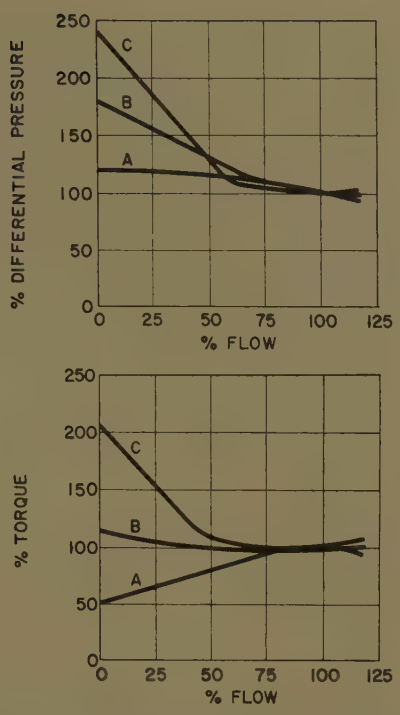


Fig. 4. Pressure-flow and torque-flow characteristics of centrifugal pumps at rated speed

A—Radial-flow pump, specific speed = 1,000
B—Mixed-flow pump, specific speed = 5,500
C—Axial-flow pump, specific speed = 10,000

An actual liquid system usually involves a combination of all three types of differential pressure. The composite system curve then appears as shown for a typical system by the solid line in Fig. 2. For simplification this paper will hereafter combine friction and velocity head into one characteristic curve which varies as the square of flow and which will be called "friction head." Specific systems will be identified by the percentage of static head compared to the total differential pressure at design flow (100% flow). For example, the composite curve *D* in Fig. 2 will be called a system with 40% static head.

Frequently the flow in a liquid system fed by a centrifugal pump is controlled by means of a variable restriction in the piping, usually a control valve. A system will have different over-all characteristics for each different position of the variable restriction, as shown by the dashed lines in Fig. 3 for a typical system with 40% static head.

When the liquid in a long full line is first accelerated by starting a pump or rapidly opening a valve, the inertia of the liquid may momentarily cause the system to appear to have a closed discharge valve.

Pump WK²

The inertia values for a variety of centrifugal pumps cannot be shown

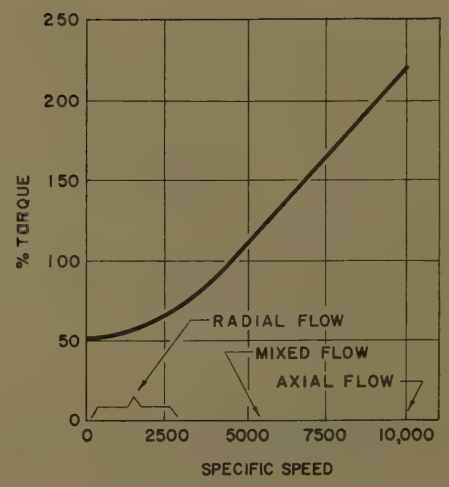


Fig. 5. No-flow torque at rated speed for centrifugal pumps having various specific speeds

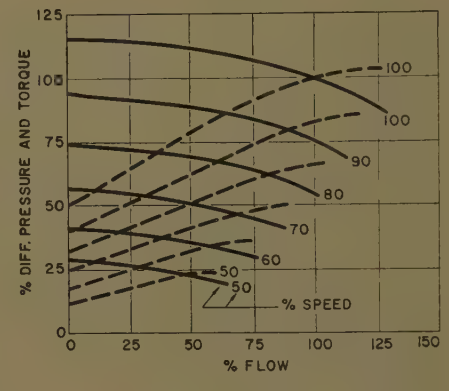


Fig. 6. Pressure and torque characteristics for a radial-flow pump operating at different speeds. Solid lines—Differential pressure. Dashed lines—Torque

conveniently by curves or tables. In general, however, the WK^2 of a motor driver is several times as great as the pump to which it is connected, and therefore accuracy in determining pump WK^2 is not usually a critical factor. Exact pump WK^2 values can be learned from the manufacturer.

Pump Characteristics

Typical constant-speed characteristics of differential pressure and torque as a function of flow are shown for three different types of pumps in Fig. 4. These three types, commonly called radial-flow, mixed-flow, and axial-flow pumps, can be classified by the term "specific speed," which is calculated as described in the Appendix.

Fig. 5 illustrates typical values of torque at zero flow and 100% speed for pumps having various specific speeds. For any individual pump, accurate characteristic curves furnished by the manufacturer show differential pressure (or head) and brake horsepower as a function of flow at design speed. Torque, which is

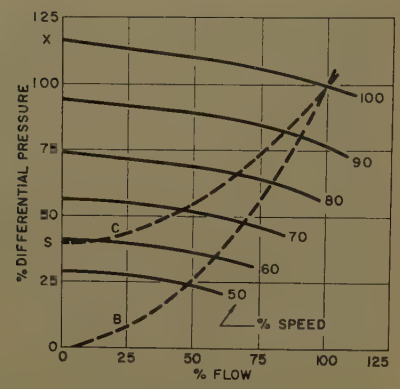


Fig. 7. Radial-flow pump and system characteristics. Solid lines—Pump differential pressure

B—System pressure with 0% static head
C—System pressure with 40% static head

directly proportional to horsepower and inversely proportional to speed, may be readily calculated for any given horsepower and speed as described in the Appendix.

To determine approximate pump characteristics at any speed other than rated speed, the affinity laws described in the Appendix may be used within reasonable limits. Typical torque and differential pressure characteristics as a function of flow for a radial-flow operating at various speeds are illustrated in Fig. 6.

Combined Pump and System Characteristics

To determine torque under any given set of conditions, the liquid system characteristic curve is plotted on the same co-ordinates as the variable-speed pump pressure-flow curves, as shown in Fig. 7. Then, for any particular speed, the corresponding flow may be determined from the intersection of the appropriate system and pump curves. If the flow is known at one particular speed, the corresponding torque may be determined from the torque-flow curve corresponding to that speed, as shown by the dashed lines in Fig. 6. By plotting a series of such torque values at several different speeds, a pump speed-torque curve may be derived.

For no-flow conditions and for flow in a system with zero per cent static head, pump torque is directly proportional to the square of speed as shown by dashed lines *A* and *B* in Fig. 8(A). However, where static head is greater than zero per cent, the curve becomes more complex, with a marked change in slope at the speed where flow begins, as shown by the solid line *C* in Fig. 8(A). In all cases described here, the pump is assumed to have a check valve in the discharge piping so that flow cannot reverse direction.

The per-unit speed where flow begins (N_c) is equal to the square root of s/x, where s is the per-cent static head of the system and x is the per cent of pump differential pressure at zero flow and 100 per cent speed. The points s and x are identified on the characteristic curves in Fig. 7. With N_c known the pump speed-torque curve may be approximated by first drawing a square relationship curve corresponding to zero flow, and then a line from speed N_c on the zero-flow curve to the 100% design point as illustrated in Fig. 8(A) for a radial-flow pump. For mixed-flow and axial-flow pumps, the resulting curve is similar to that illustrated in Fig. 8(B).

In certain applications, a control valve or other variable restriction in the fluid system will change position at the same time that pump speed is changing. For example, a control valve instrumented to maintain constant flow will open as the pump slows down, in attempting to maintain the desired flow rate. Different positions of a variable restriction produce a family of system characteristic curves as shown by the dashed lines in Fig. 3, wherein each system curve is identified by the corresponding flow at 100% pump speed.

Fig. 9(A) illustrates how the different positions of a variable restriction affect the speed-torque characteristics of a radial-flow pump working in a system with 40% static head. By plotting known torque values at 100% speed, these characteristics may be approximated by eye, since all curves coincide at N_c, the speed where flow begins.

Pump Torque Characteristics During Transient Conditions

Most radial-flow centrifugal pumps are normally started with the discharge valve closed, resulting in a pump speed-torque characteristic which varies as the square of speed and which has a full speed-torque value as determined by Fig. 5. Mixed-flow and axial-flow pumps are usually started with the discharge valve open in order to minimize the starting torque required. To determine sta
ing speed-torque characteristics of th
pumps when static head is greater th
zero it is necessary to plot the system a
pump curves as described above. T
inertia of the liquid in a long full line m
increase the starting torque requireme
of mixed-flow and axial-flow pumps
making the system appear to have

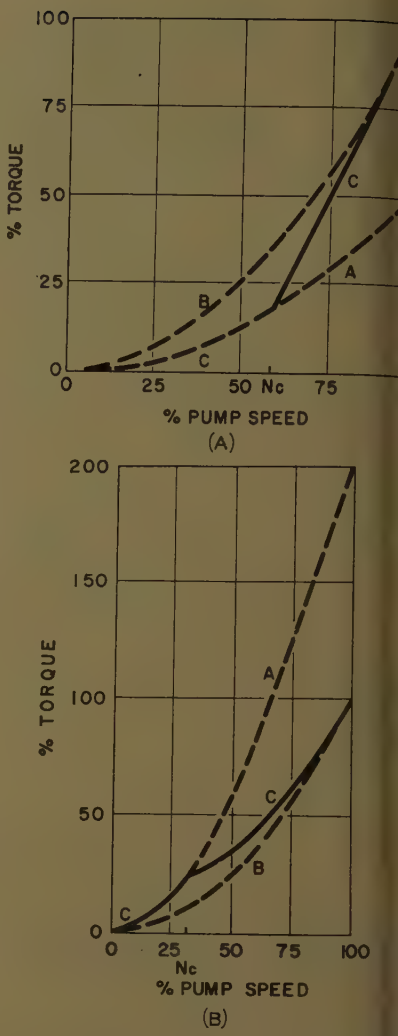


Fig. 8. Speed-torque characteristics. (A for a radial-flow pump. (B)—For an a flow pump

A—No-flow curve
B—Flow with 0% static head
C—Flow with 40% static head

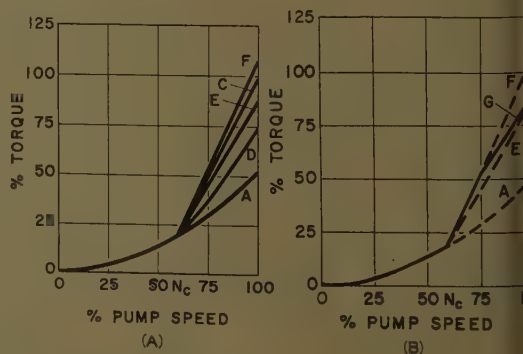


Fig. 9. (A)—Speed-torque characteristics for radial-flow pump operating in various systems. (B)—During deceleration. System flow as follows: A—No flow. C—100%; D—50%; E—75%; F—125%; G—When control valve is opening

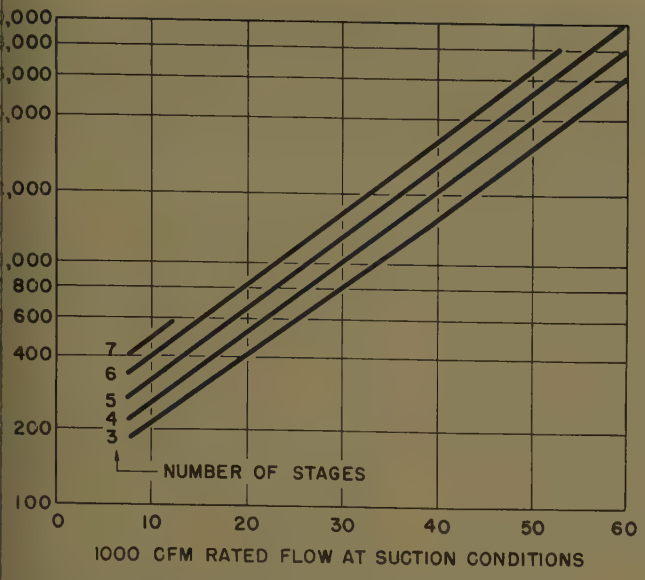
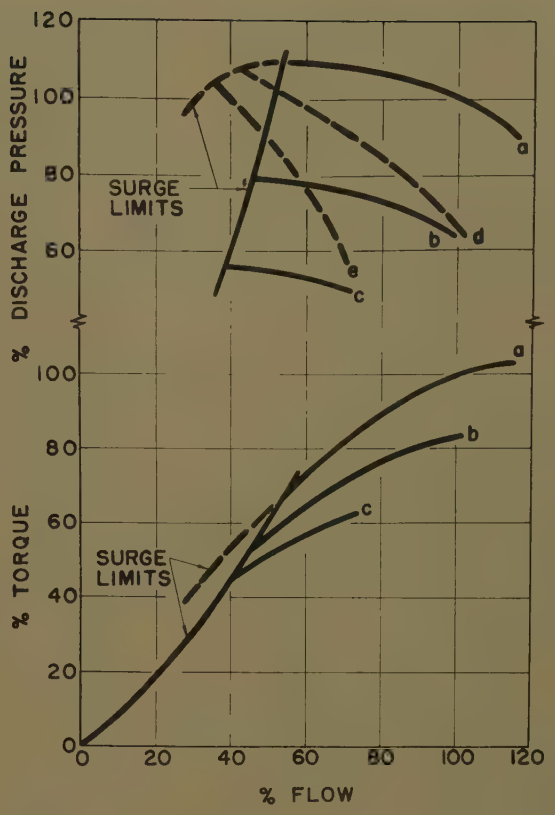


Fig. 10. Typical WK2 values for centrifugal compressors



11 (right). Centrifugal compressor pressure and torque characteris- Solid lines—Unthrottled characteristics at various speeds: a— %; b—90%; c—80%. Dashed lines—Throttled characteristics at 100% speed

ed or partially closed discharge
ve.

he calculation of speed-torque char-
eristics of all types of pumps during
nentary power failures may require
ting the system and pump curves.
example, assume the case of a radial
pump and a system whose character-
s are illustrated by Fig. 6 and Fig. 3.
ume that, under normal conditions,
system curve is as shown for 75% of
design flow, and that, when electric
er fails, the control valve opens wide
known length of time, changing the
em curve to that shown as 125% of
gn flow. The speed-torque curve for
decelerating phase of the momentary
re will be as shown by the solid line
Fig. 9(B), the portion of the curve
tified as *G* representing the period
the control valve is moving from the
% position to the 125% position.
en the motor is re-energized, the
p speed-torque curve will be as shown
125% of design flow, since the control
e will likely remain fully open until
machine has returned to 100% speed.

mpressors and Gas Systems

System Characteristics

wo types of gas systems are com-
ly associated with centrifugal com-
sors in refineries. In a simple com-
sion system, suction and discharge
sures remain nearly constant for var-
flow rates, since only a small portion
he differential pressure results from
ion. In a circulating system, suction pressure usually remains constant but differential pressure varies approximately as the square of flow under normal operation. During transient conditions the discharge pressure in a gas system will not change as rapidly as will pressure in a liquid system, the rate of change being a function of the system volume and the means by which system pressure is maintained. For example, the gas systems associated with catalytic-cracking-unit main air blowers and gas-concentration-unit compressors have relatively large capacities and are operated at comparatively constant suction and discharge pressures. In these systems, pressure changes are relatively small during most transient conditions. A constrasting system is that exemplified by a recycle gas compressor in a catalytic reforming unit where the relative system capacity is low and there is no control equipment attempting to maintain a fixed pressure at the discharge of the compressor. In this type of recirculating system, pressure change can be significant within a few seconds. For the purpose of this paper, gas systems will be described only as to whether the compressor discharge pressure changes slowly or rapidly during transient conditions.

Compressor WK^2

Typical values of WK^2 for centrifugal compressors are shown in Fig. 10. Since most motor-driven centrifugal compressors are connected to the driver through a speed increaser, the effective inertia of a compressor referred to the motor shaft must be determined by multiplying the compressor WK^2 by the square of the ratio of compressor speed to motor speed. To obtain total load WK^2 the inertia of the gear unit and couplings (referred to the motor shaft) must be added to the effective inertia of the compressor.

Compressor Characteristics During Steady-State Conditions

Typical pressure-flow and torque-flow performance curves for a centrifugal compressor with fixed suction conditions are shown in Fig. 11. The solid lines represent unthrottled performance at various speeds *a*, *b*, and *c*. The characteristics are discontinued for flow less than approximately 50% of design flow because unstable or surging operation occurs in this range. With suction throttling, such as the use of inlet guide vanes, the operating range can be extended, as shown by the dashed lines. Lines *d* and *e* represent characteristics based on two different positions of the guide vanes at a constant compressor speed corresponding to curve *a*. Normally, centrifugal compressor performance curves for various speeds may be obtained from the manufacturer. If only the rated-speed curves are available, the unthrottled pressure and torque characteristics at other than rated speeds can be approximated by using the relationships described in the section "Com-

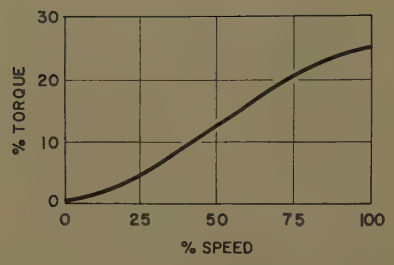


Fig. 12. Starting torque characteristics with closed guide vanes

pressor Characteristics at Various Speeds," in the Appendix.

The discharge pressure developed by a compressor at constant flow and constant speed varies widely, depending upon the nature of the gas at suction conditions. Therefore, in analyzing transient compressor behavior as it affects the motor driver, it is best to work from performance curves pertaining to that particular gas and set of conditions which impose the greatest anticipated load on the driver. For example, the greatest load on an unthrottled air compressor with atmospheric suction will occur during extremely cold temperatures. The compressor manufacturer considers 100% torque as that corresponding to design horsepower at rated speed and at design suction conditions which, of course, has no direct relationship to the name-plate rating on the motor driver.

Characteristics During Transient Conditions

The normal starting of a centrifugal compressor seldom presents a torque problem. With the compressor suction valve or inlet guide vanes closed and with the discharge system being at nearly the same pressure as the suction system, the compressor torque at full speed can usually be kept to less than 25% of full-load torque, with torque at lower speeds varying roughly in proportion to the speed, as shown in Fig. 12. However, an analysis of the torque requirements during momentary power failures usually demands a more rigorous investigation, as described by the following examples.

Fig. 13 describes pressure and torque at various speeds plotted against flow of an unthrottled compressor. Assume that the unit is operating at its rating, point *A*, when the voltage fails to the motor driver. If the pressure remains relatively constant during the decelerating period, the compressor operation will follow the path identified by line *AB*. At *B* the surge range is entered, and soon after passing the surge limit, the compressor will not be able to develop enough discharge pressure to force gas into the system. The discharge check valve will therefore close and cause the compressor to operate at a no-flow condition. Torque characteristics at no-flow and elsewhere in the surge range are subject to some speculation since prolonged testing may mechanically damage the compressor, and the rapidly fluctuating conditions are difficult to record with conventional instrumentation. To simplify, it may be assumed that torque at the surge line is identical to that at no flow, for any one speed. It may be further assumed that no-flow torque varies as the square of the speed, with the full-speed torque being that shown as point *E*. On this basis the no-flow speed-torque curve will be shown by solid line *EBCO* in Fig. 14. The actual torque values probably lie somewhat below the curve as shown, but conservatism and lack of accurate test data suggest using this plot. Referring Fig. 14, upon voltage failure, the sp drops quickly from operating point to no-flow point *B*, and then more more slowly toward complete rest at p *O*.

If the gas system is such that the charge pressure decreases rapidly w the machine decelerates, the operatio an unthrottled compressor will follo path similar to line *AC* in Fig. 13. In case the compressor will decelerate t lower speed before entering the su range, and the corresponding spe torque curve assumes the shape shown line *ACO* in Fig. 14. A situation wh discharge pressure decreases during transient period may require a trial error or step-by-step solution in orde establish the rate of change in c pressor speed and system discharge p sure, since these two factors may mutually dependent.

When the motor driving a centrifu compressor is re-energized following momentary voltage interruption, compressor speed-torque curve dur the accelerating period will be similar that described for the decelerating per if discharge pressure has not chang The curve will differ, however, if a s stantial decrease in discharge press takes place during the momentary volt failure. Dashed line *CD* in Fig. 13 sh the pressure-flow plot assuming that discharge pressure has decreased to 9 of normal and will remain at this le during the entire reaccelerating peri At 90% pressure, flow at full speed increase to approximately 115%. *OCD* in Fig. 14 illustrates the result speed-torque characteristics for the celerating period.

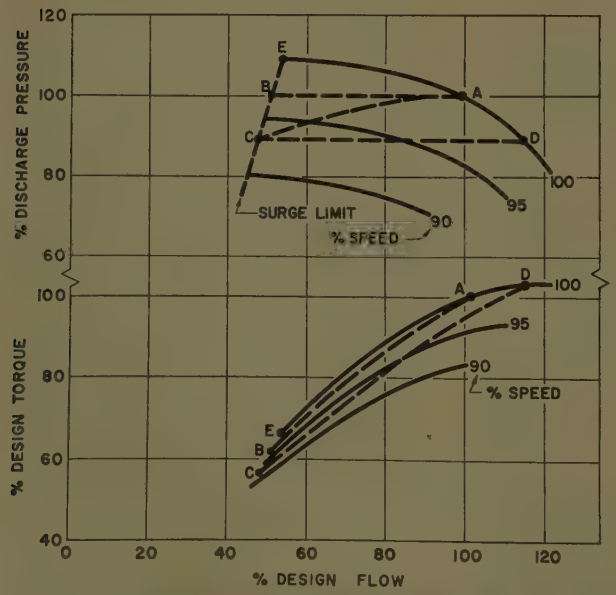


Fig. 13. Unthrottled compressor characteristics at various speeds

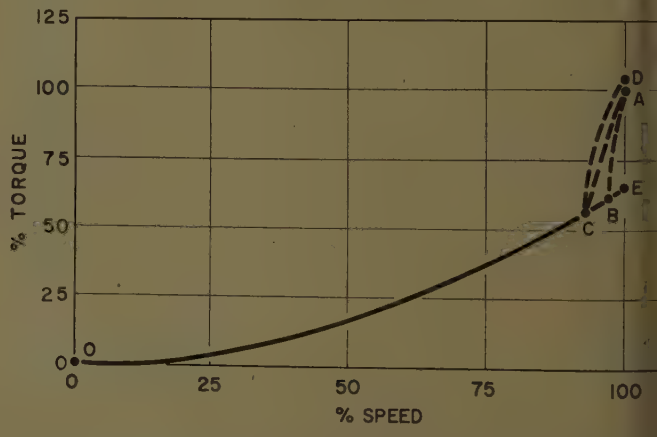
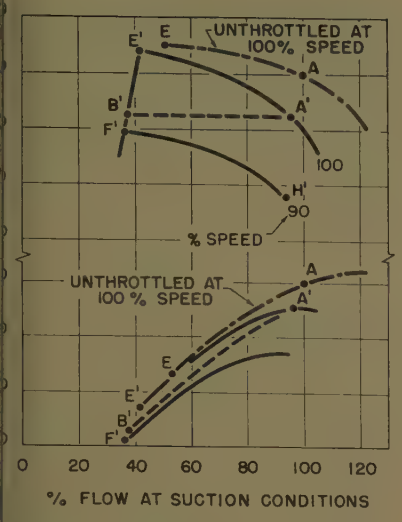


Fig. 14. Unthrottled compressor speed-torque characteristics various system conditions. Discharge system pressure as follow

EBCO—No-flow
ABCO—100%
ACO—Decreasing during deceleration
DCO—Constant at 90%



15. Compressor characteristics at various
ds for one typical position of suction
throttling device

RACTERISTICS WITH SUCTION
HROTTLING

many constant-speed compressor
ications, the flow is controlled by
ns of a damper, valve, or set of guide
es located on the suction side of the
pressor. The effect of such a variable
riction is the lowering of gas pressure
the compressor intake which, in
, reduces the weight flow of gas and
discharge pressures. A suction-
ttling device affects the speed-torque
acteristics of a centrigual compressor
ng transient conditions in various
s, depending upon its speed of travel
how it is arranged for operation dur-
power failures.

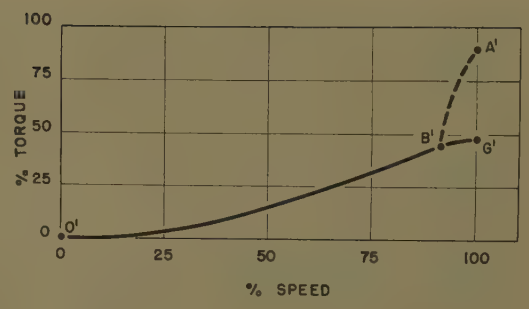
electric power fails to the compressor
or, but the instruments and auto-
ic operator which control the suction-
ttling device remain operative, the
ce will probably open wider when
compressor decelerates, attempting
naintain constant gas flow. When
throttling device is completely open,
compressor characteristics will be as
ribed previously for unthrottled ma-
es.

the suction-throttling device main-
s its position when electric power
, or if it moves so slowly that there
nly slight change in position during
transient period, the compressor
acteristics will be similar to those
vn in Fig. 15. The usual unthrottled
acteristics are shown as dash-dot
. Instead of operating at the design
t A, the compressor is assumed to be
ially throttled in order to run at
t A'. Because of the large amount
rottling the comressor is capable of
le operation at flows less than those
ed by the normal or unthrottled

Fig. 16. Throttled compressor speed-torque characteristics. Discharge system pressure as follows:

O'B'A'—Constant at 85%, partial throttling

O'B'G'—Constant, closed throttling device



surge limit. Pressure-flow curve $E'A'$ estimates the full-speed performance of the unit with the throttling device in one position. Line $F'H'$ represents the performance at 90% speed with the throttling device in the same position. Line $E'B'F'$ traces the corresponding surge line at various speeds. Upon momentary voltage failure to the motor driver, the compressor loses speed rapidly and, assuming a discharge pressure which remains relatively unchanged, the flow decreases from A' to B' as speed drops from 100% to 92%. The torque during this transition period is represented by dashed line $A'B'$ in Fig. 16, and the over-all speed-torque curve assumes the shape shown as line $A'B'O'$.

In some installations, the compressor suction-throttling device will be instrumented to close automatically whenever voltage fails, in order to unload the machine during the transient period. Under these conditions the speed-torque requirements during deceleration depend upon the speed at which the throttling device closes. In many installations the throttling device will not move significantly before the compressor has decelerated to the speed at which it can no longer overcome the system pressure and the speed-torque characteristics will be as described in the preceding paragraph. Using this assumption will produce conservative results. During reacceleration the compressor torque will follow the square relationship no-flow curve until the speed is reached where flow begins. If the guide vanes have closed nearly all the way by the time flow is re-established the resulting speed-torque curve will be similar to that shown by line $O'B'G'$ in Fig. 16. Torque at full speed is approximately equal to the torque occurring at the full-speed throttled surge limit corresponding to the existing system discharge pressure.

Conclusions

Determination of the torque requirements of centrifugal pumps and compressors during transient conditions requires detailed analyses whenever the machines operate into discharge systems containing some static pressure. Under such conditions, torque varies as the square of speed only in the no-flow range. In the flow range, speed-torque characteristics assume a complex shape which must be determined from the machine and system characteristic curves.

Appendix. Calculations

Specific Speed

Specific speed for any pump may be calculated by

$$N_s = \frac{NQ^{1/2}}{H^{3/4}}$$

where

N_s = specific speed
N = rated pump speed, rpm
Q = flow at maximum efficiency, gallons per minute
H = head per stage at maximum efficiency, feet

Torque

Torque may be calculated by the following expression when both horsepower and speed are known:

$$T = \frac{HP \times 5{,}250}{N}$$

where

T = torque, pound-feet
HP = horsepower
N = speed, rpm

Pump Characteristics at Various Speeds

The affinity laws by which approximate pump characteristics may be determined at speeds other than the rated speed are as follows:

Flow F varies directly with speed.
Head H, or differential pressure, varies as the square of speed.
Horsepower HP varies as the cube of speed.
Torque T varies as the square of speed.

The above laws are applied by first selecting a point F_1H_1, on the pressure-flow curve at rated-speed N_1. The corresponding point F_2, H_2 on the pressure-flow curve for the new speed N_2 is determined by multiplying F_1 by N_2/N_1 and by multiplying H_1 by $(N_2/N_1)^2$. The HP_2 at N_2, and $F_2 = HP_1 \times$

$(N_2/N_1)^3$. T_2 at N_2 and $F_2 = T_1 \times (N_2/N_1)^2$. After plotting several different points in this manner, continuous curves of pressure flow, horsepower flow, and torque flow can be drawn for the new N_2.

Compressor Characteristics at Various Speeds

The affinity laws for centrifugal compressors are identical to those described above for centrifugal pumps except that the compressible nature of gas complicates the calculation of differential pressure. Usually it is convenient to think of compressor performance in terms of compression ratio, i.e., the ratio of absolute discharge pressure to absolute suction pressure. When gas conditions at the compressor inlet are constant, compression ratio at any speed N_2 may be determined by

$$r_2 = \sqrt[\frac{k-1}{k}]{(N_2/N_1)^2\left(r_1^{\frac{k-1}{k}} - 1\right) + 1}$$

where

r_2 = compression ratio at N_2
k = ratio of specific heats for the particular gas being compressed (1.4 for air)
N_2 = new speed
N_1 = design speed
r_1 = compression ratio at N_1 for the particular point considered

When r_2 is determined, the corresponding absolute discharge pressure P_2 is calculated by multiplying the suction pressure by compression ratio. This P_2 occurs at which is determined by multiplying by N_2/N_1. Plotting a series of such p will define a pressure-flow character curve at N_2. Horsepower HP_2 at equal to HP_1 multiplied by $(N_2/$ T_2 at F_2 is equal to T_1 multiplied $(N_2/N_1)^2$.

References

1. Centrifugal and Axial Flow Pumps (b A. J. Stepanoff. John Wiley & Sons, Inc., York, N. Y., 1957, pp. 285–92.

2. Centrifugal Pumps and Blowers (b Austin H. Church. John Wiley & Sons, Inc., pp. 189–216.

Discussion

H. C. Barnes and **A. F. Gabrielle** (American Electric Power Service Corporation, New York, N.Y.): The modern refining operation is highly electrified and requires continuity both of the electrical supply and of the many pieces of rotating equipment which constitute the continuous process. Continuity of electrical supply is seldom a problem as a system of parallel feeds with circuit-breaker switching usually can be worked out on a basis which is economically fair to both the utility and the user.

Not so readily solved is the problem of maintaining stability of the motor-driven equipment during and after switching of transient disturbances on both the utility and users systems. It is most discouraging to provide good 2-way supply and then to have stability lost on simple faults on the utility or users system because proper consideration was not given to the rotating equipment. Motor and mechanical equipment characteristics, transformers and system impedances, and synchronous motor excitation methods are among the many items contributing to stability. Unless they are properly considered early enough to assure proper design, even the highest speed relaying and circuit breakers may not be able to clear fast enough to prevent instability. Fortunately, the problems are not insurmountable if properly analyzed in the preliminary design stage; in fact, the remedies are quite often simple and inexpensive, but not if postponed until the installation is completed.

To aid the customers of the American Electric Power Company System (which includes the following companies: Appalachian Power, Indiana and Michigan Electric, Kentucky Power, Kingsport Utilities, Ohio Power, and Wheeling Electric) to attain the proper design at an early stage, we developed a digital computer program for determining critical switching times. This program and the necessary supervision are available free of charge to our customers; the program is also available at no charge to others with only the stipulation that they demonstrate the ability to put it to proper use.

One of the difficult tasks in these studies is the obtaining of data on the customers' system, particularly in regard to the speed-torque characteristics of the mechanical load connected to the motor shaft. It appears from this paper that, for the purpose of determining critical switching time, the load torque of compressors and pumps can be assumed to vary as the square of the speed during electrical transients. This assumption makes computation quite simple and is in the direction of providing a margin of safety.

The operation of check-valves at various stages of acceleration and deceleration stimulates speculation as to its possible increased use to assist reacceleration of equipment.

The authors have made a real contribution in analyzing and presenting their data. It is hoped that others will be encouraged to make similar analyses of other types of driven equipment.

C. A. Johnson and **A. R. Kelly** (Esso Research and Engineering Company, Madison, N. J.): This paper is heartily welcomed. Not only does it tie together, for the first time we know of, the centrifugal machine and system characteristics which are of interest to the electrical engineer; it also shows how these characteristics can be manipulated to give, for different transient hydraulic conditions, the dynamic speed-torque characteristic of the machine in question on the system in question. Granted the availability of the required pressure- and torque-flow curves, there is no longer an excuse for the oversimplified assumptions, so common in the past. For most radial-flow pump and compressor applications, these assumptions have led to overpessimistic conclusions.

It is noted that all discussion of compressors, including the affinity laws given in the Appendix, is based on maintaining suction conditions constant at those for which the curves are drawn. The applicability of this paper would be enlarged by a discussion of the effects of rapidly changing suction conditions. An example is the compressor of a refrigeration system, in which discharge pressure remains constant but suction pressure rises rapidly, following loss of compression.

It may be difficult to apply the principles given for compressors with throttled suction. The American Petroleum Institute tentative standard *617* specifies the performance curves to be included in the proposal. As for characteristics under load of motor-driven machines, these consist of a single pressure-flow and horsepower-flow curve, based on normal suction pressure. For a machine with butterfly-valve throttling, with flow plotted in actual cubic feet per minute, at the eye of the first-stage impe the curves will represent a variable va position case, where valve position va with flow, to keep machine suction pres constant at the design point. For su case, it would presumably be necessar request additional curves from the pressor manufacturer, perhaps based o suction pressure expected ahead of butterfly valve.

Along the same lines, further clarifica might be desirable regarding the s shown in Figs. 11 and 15, of surge limit suction throttling. Again, assuming plotted in actual cubic feet per minute, understood that the throttled charact tics shown apply to a machine with guide vanes, which act to alter compres ratio.

For butterfly-valve suction throttling the other hand, the surge limit is unders to remain at the same flow for a given sp the pressure characteristic drops with creased throttling and, if plotted for stant valve position, this would become drooping with increased flow due to incre pressure drop across the valve.

Additional clarification would be he regarding the shape and magnitude of pressor no-flow speed-torque curves. in the section which discusses character during transient conditions, it is said torque varies about in direct proportio speed, as indicated by Fig. 12, for a no closed-suction start, reaching perhaps 25% full-load torque at full speed. pressured discharge is assumed. Se in the same section, it is said that th flow torque in the surge range varies a square of speed during deceleration, w full-speed value somewhat below the to at the full-speed surge limit (about 65 the illustration); the reaccelerating s torque curve during the no-flow phase is to be similar. Open suction and blocked pressured) discharge are assumed. T in the following section, it is reiterated unthrottled no-flow acceleration ag closed discharge results in square rela ship of torque to speed. For fully suc throttled reacceleration, the illust speed-torque curve of Fig. 16 is roug square relationship, except with a re bend above 75% speed, and reaching n 50% torque.

It would be helpful if the authors v explain why the no-flow torque is diff in the first case from that in the other as well the reason for difference betwee flow torques in the latter two cases.

he Application of Static Switching to e Control of Two 7,500-Hp Oil-Fired Combustion Turbines

P. T. CARMACK
ASSOCIATE MEMBER AIEE

E. M. SMITH
MEMBER AIEE

ITCHING of electric power is one of the most basic problems of the rical engineer. The most common accepted method of switching is by hanically joining and separating ucting parts. During the past 10 s, the art of switching electric power cally has been developed to the point it can now compete with the me- ical types of switching devices, par- arly in the control field.

he application presented in this paper new one. It is hoped that this entation will contribute toward the her development of the art by en- aging closer study of applications of c control which may not have been idered feasible heretofore. Further rmation on this subject may be d in the references.[1–8]

cription of the Installation

his installation of two gas turbines cated in a chemical plant of Union oide Caribe Inc. at Ponce, Puerto . The plant site is on the Caribbean coast, but there is very little rainfall pt for 3 months of the year.

s shown in Fig. 1, each turbine is to drive a mechanical load consist- of an air compressor and is rated 0 horsepower. The turbines are of single-shaft simple-cycle, nonregenera- type. It is necessary to rotate the ines with an external source of power re they can be fired. This is ac- plished by the small cranking steam ines. Three pumps are provided he lubrication system, which provides under pressure for operating the rnor and a hydraulic control mech- n in addition to providing lubrication the turbines and driven machine. The three pumps are a main-shaft-driven pump, an auxiliary motor-driven pump, and an emergency d-c motor-driven pump.

Two shaft-driven fuel pumps are provided for each machine. Either distillate (diesel) or residual (bunker *C*) fuel is used at the operator's discretion. However, it is necessary to start and stop the machines on distillate fuel.

Combustion fuel is atomized. A shaft-driven atomizing air compressor is provided on each turbine and a motor-driven starting air compressor is common to both machines.

Exhaust gases from the turbines contain a high percentage of air at high temperature. Heat is extracted from the gases in a waste heat boiler. The exhaust of the waste heat boiler may be directed to the intake of an oil-fired boiler to improve efficiency. The boilers are used to produce process steam.

Gas turbines were selected as drivers for the air compressors in preference to other types of drivers, such as electric motors, steam turbines, or diesel engines, as a result of a study which showed these turbines to be most economical. High cost of distillate fuel and electric power were contributing factors toward the selection. Heat recovery from the exhaust gases was not considered in the initial study, but its addition further justified the selection.

A view of the utilities area during the latter stages of construction is shown in Fig. 2. Turbine No. 2 is located at the left center and turbine No. 1 at the right center of the area. Fig. 3 is a close-up of turbine No. 2 showing the compressor and the turbine shell with combustion chambers removed.

The control equipment for both turbines is contained in a 3-unit duplex-type switchboard; see Fig. 4. The two end units contain duplicate equipment associated with the individual machines while the middle unit contains equipment associated with both machines. Fig. 5 shows an interior view of this duplex control switchboard.

Why Static Control?

The control system for this application was derived from a design which was developed primarily for isolated locations where a minimum of experienced operators would be attending the unit. Experience of other users of gas turbines installed in unfavorable atmospheric conditions had shown that one of the chief causes of incorrect functioning of the equipment had been caused by failure of contacts or of other elements of electromechanical devices. Two methods of decreasing down time caused by malfunction of the control system were considered.

The first method considered was the elimination of most of the automatic control devices. This, of course, meant manual control. During the early stages of design, the capability of the turbine operators was an unknown factor. Therefore, it was assumed that they would be for the most part inexperienced. Since the operation of the whole plant depended on the turbines, the idea of adding to the operators' responsibilities was discarded.

The second method considered was static control. The reliability of this type of control is inherent in the fact that there are no moving parts. Another feature which adds to its reliability is that the elements are encapsulated in a material which is impervious to moisture

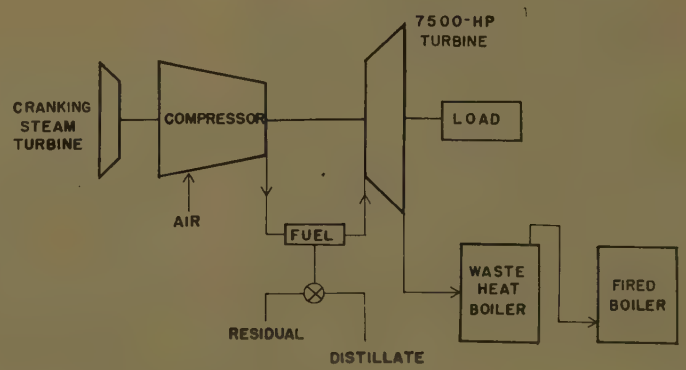


Fig. 1. Block diagram showing arrangement of machines and process steam boilers for one turbine

r 60-52, recommended by the AIEE Petroleum stry Committee and approved by the AIEE nical Operations Department for presentation e AIEE Winter General Meeting, New York, , January 31–February 5, 1960. Manuscript itted November 2, 1959; made available for ing December 10, 1959.

. CARMACK is with the Union Carbide Inter- nal Company, Houston, Tex. E. M. SMITH th the General Electric Company, Philadel- Pa.

Fig. 2. Utilities area during latter stages of construction

Fig. 3. Close-up of one turbine and its compressor during assem

This is a very desirable feature in this case because the atmosphere is not only corrosive, because of the tropical climate and the close vicinity of the sea, but it is also very dusty because of the long dry periods.

Static switching has been tried and proved in the machine tool industry where highly repetitive operation is required. However, in this case, it was selected not so much for repetitive operations as for the fact that it would be located in a corrosive and dusty atmosphere.

The one objection to the use of static control is the higher cost. In this case, the cost was further increased by the requirement for operating the system on either d-c or a-c control power. This was necessary because the plant was designed to operate, for a limited period, independent of contract a-c power.

The additional cost was justified in this case on the basis that two days lost production would offset it.

Operating Requirements

As previously pointed out, the high cost of distillate fuel made it desirable to operate the machines from residual fuel. However, since distillate fuel is required for starting and stopping the machines, the control was arranged so that, at the operator's discretion, the units could be operated on either fuel. During starting, the distillate fuel provides a means of bringing the machines up to operating speed and temperature where the residual fuel can be successfully used. On stopping, it is used to purge the machines of the heavy fuel thus preventing residue from forming when the turbines cool.

Since it was desired that the machines operate under the supervision of an operator, yet require a minimum of his attention, a partially automatic type of control was provided. With this arrangement the operator's functions were limited to the following:

1. Selection of a-c or d-c control power.
2. Start and stop the turbine auxiliaries, i.e., the lubricating pumps, air compressor, fuel pumps, etc.
3. Open and close the necessary valves, i.e., the cooling water, fuel, and distillate lubricating supply valves.
4. Starting and stopping the cranking turbine, including engaging the clutch to the main turbine.
5. Operate the control switches and push-buttons to cause the turbines to start, run, stop, or shut down in an emergency.
6. Initiate transfer from distillate to residual fuel and back to distillate. This transfer of fuel is automatically monitored to prevent an operation which would be damaging to the turbine.

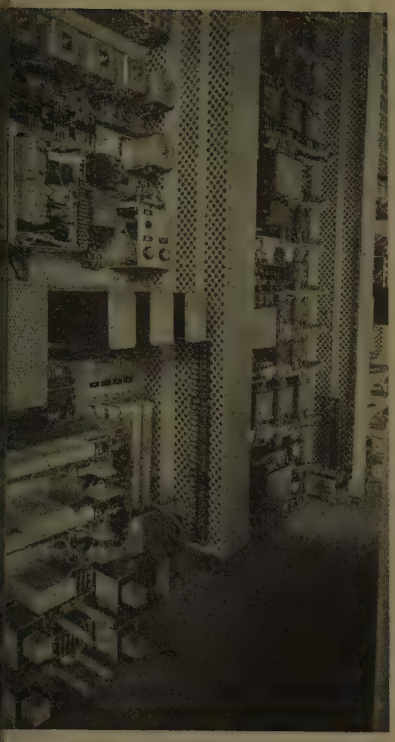
The controls are arranged to provide protection of the turbines and to shut them down automatically should the following abnormal conditions occur:

1. Turbine overspeed.
2. Loss of distillate lubricating oil pressure.
3. Loss of flame.
4. High turbine exhaust temperature.
5. Operation of protective devices wh monitor the process.

Comparison Between Electromechanical and Static Control Circuits

Before laying out the more comm circuits using electromechanical rela the designer needs a thorough knowle of the characteristics of the devices intends to use. Such knowledge is als prime requisite when using static co ponents. Basically, the two syste differ in that with relays a single input a coil is converted into various outp by means of a moving armature's clos or opening its contacts. On the ot hand, with the usual static control co ponents, a combination of inputs contr a single output. In this particular stallation static components of the s saturating magnetic amplifier type w used. A combination of inputs is t converted into an output by controll

Fig. 4. Front view of duplex control board for the two turbines



. Static control devices inside of duplex switchboard

mpedance of an output coil which is nd on a special magnetic core. As be seen from the following comparison difference leads to procedures which somewhat foreign to individuals stomed to dealing with electro-hanical relays and their circuitry.

order to assist in understanding the ation of control equipment which zes relays, it is customary to pre- a schematic or elementary diagram how, progressively, electrical con- ions between device coils and con- . It is not possible to prepare such agram when static components are , due to lack of contacts. There has efore been adopted the so-called gic Diagram" for use with static con- Such a diagram shows the flow of ts and outputs between the various ponents together with symbols which ate how they function. The sym- are shown in Table I. The device gnations and abbreviations used on various drawings contained in this r are listed in the Nomenclature.

order to illustrate more fully the rence between these two types of rol a portion of the system com- ng a common timing circuit and ting chain has been chosen for the ose of making a detailed comparison. portion is shown inside the block ed by the dotted lines in Fig. 10. is common timing circuit and count- hain is used for start-up purge, start- turbine acceleration, and shutdown

Table I. Device Symbols

Component	Component Function	Diagram Symbol
Two Input AND	Will produce an output only when both inputs are energized	
Four Input AND	Will produce an output only when all four inputs are energized	
Three Input OR	Will produce an output when any one (or more) input is energized	
Two Input NOT	Will produce an output only when both inputs are not energized	
Release Memory	Will remember which input was last energized and will produce a corresponding output Will revert to an OFF state upon interruption and restoration of power	(ON) (OFF)
Adjustable Time Delay	Will produce an output at a preset time interval following application of continuous input	
Signal Converter (D-c original input)	Will convert 115-volt filtered d-c to proper logic element operating level when energized by pilot device	
Signal Converter (A-c original input)	Will convert 115-volt 60-cycle a-c to proper logic element operating level when energized by pilot device	

Component	Diagram Symbol	Component	Diagram Symbol
Amplifier		Resistor	
Transistor		Capacitor	
Rectifier		Transformer or Reactor Winding	
Controlled Rectifier		Fuse	

purge. Essentially the circuits consist of three major systems as follows:

1. The discriminating circuits which have the ability to select a particular timing cycle and feed the "cycle complete" signal to the proper sequence controls at the completion of the timing cycle. Each of the discriminating circuits has its own timing element by which the time cycle can be varied independently of the other cycles.

2. A common timing circuit which performs the basic timing function and to which the independent timing elements are added.

3. A common "counting chain" which accumulates a predetermined number of timing impulses to obtain a timing period much greater than available from any single static timing device.

This system greatly minimizes the total number of actual timing elements necessary to obtain the relatively long increments of time required.

The common timing circuit consists of time-delay elements TD4 and TD5 and associated elements 1A12, 1A13, 1A17, 2A2, 2A3, 2A4, and 3A15. The common timers TD4 and TD5 have a fixed timing cycle of approximately 20 seconds each. The independent timers TD1, TD2, and TD3 may be adjusted to make slight changes in complete time cycles.

The common counting chain consists of 10 Release Memory logic elements 2A8 through 2A17 and OR element 1A12. The Release Memory elements are connected to obtain an OFF override function. That is, the OFF signal is always present and two ON signals are required to override the OFF signal and obtain an output signal from the unit. Major changes in the total timing cycles may be made by changing the point in the counting chain where the "timing cycle complete" signal is obtained.

In order to explain the operation of these circuits, assume the turbine has been given the shutdown signal while burning bunker *C* fuel and the fuel transfer to distillate fuel is complete, all of the release memory elements in the counting chain have an OFF signal applied and the timing circuits have all been reset from previous operations.

The discriminating circuit, from F in Fig. 9, provides an input signal to element 2A6 and 1A12. The signal to 1A12 causes it to turn on and supply an input

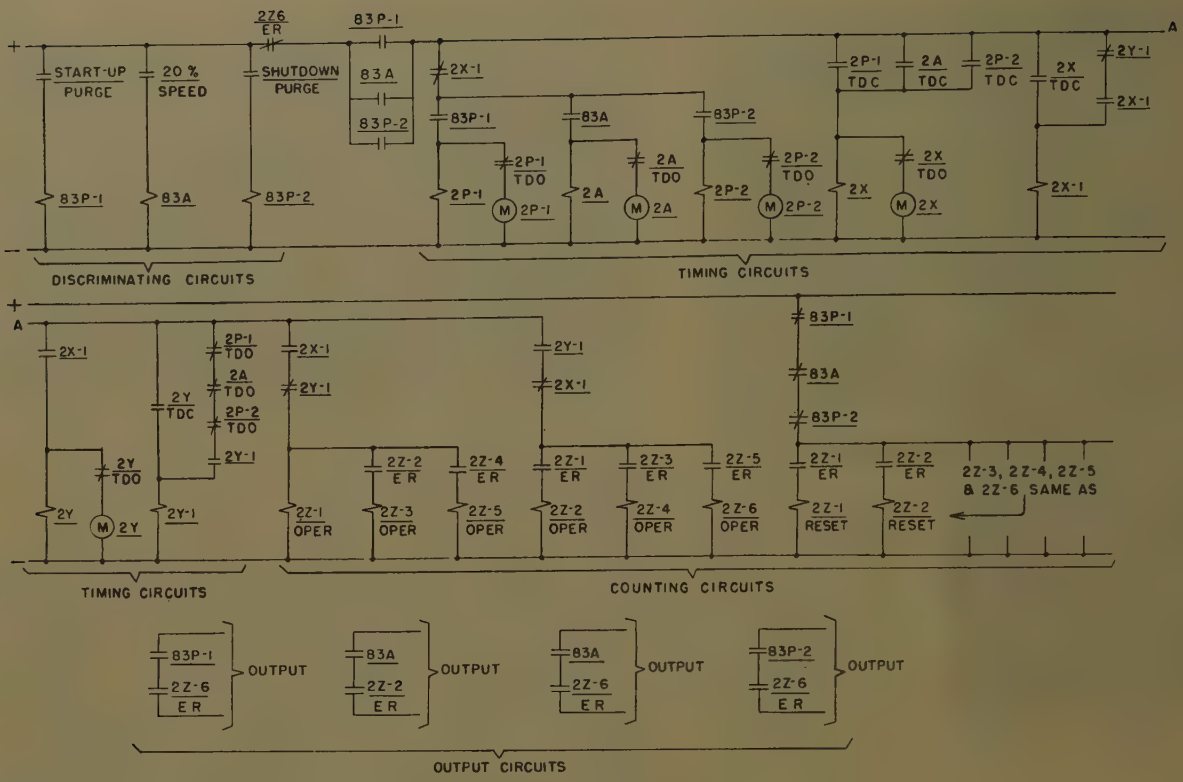


Fig. 6. Ele
tary dia
showing
circuits
counting
using ele
mechanical re

signal to element 2A8 and element 2A4. Counting chain element 2A8 does not turn on because the OFF signal is still present and the single ON signal is not enough to override it. However, element 2A4 does turn on since it already has one signal from NOT element 2A3. The output signal from 2A4 causes element 1A13 to turn off, resetting timer TD4. In addition, the signal from 2A4 starts timer TD5 through element 3A15. When TD5 times out, it applies an input signal to each of elements 2A5, 2A6, and 2A7. However, only element 2A6 turns on since it already has a signal from an element of the discriminating circuit and the other two elements have only the one signal from TD5. Element 2A6 causes TD2 to sequence. The signal from 1A17 causes three things to happen:

1. It causes elements 2A3, 2A4, and 3A15 to turn off in sequence and thereby reset timer TD5; in addition, element 1A13 is turned on.

2. Element 2A2 turns on (it already has a signal from 1A12) and applies an ON signal to counting chain element 2A8, 2A10, 2A12, 2A14, and 2A16. Element 2A8 is the only one of these to turn on since it is the only one to have two input ON signals

3. When element 1A13 was turned on, as a resulting of losing its input signal from element 2A4, it started timing element TD4.

At the completion of its timing cycle, TD4 comes on to cause element 1A14 to turn on and element 1A17 to turn off. The removal of the input signal from element 1A17 to element 2A2 causes it to turn off and remove an ON signal from elements 2A8, 2A10, 2A12, 2A14, and 2A16. Element 2A8 remains on, however, since the OFF current is not sufficient to override the remaining ON signal from 1A12. At the same time, element 2A3 turns on, having lost an input signal from element 1A17. With element 2A3 having an output signal, element 2A14 will come on again. This time, the output of 2A4 causes three operations to occur simultaneously as follows:

1. An ON signal is applied to elements 2A9, 2A11, 2A13, and 2A17, resulting in element 2A9 turning on since it already has one ON signal from element 2A8.

2. Timer TD4 is reset as the result of element 1A13 turning off when it received an ON signal from element 2A4.

3. Element 3A15 turns on to start timer TD5 timing for a second time.

This action continues until counting chain element 2A17 receives an ON signal from element 2A16 and an ON signal from element 2A4, causing it to turn on. The output signal from 2A17 is used to cause the turbine to shut down by operating the hydraulic dump valve solenoid device 20 HD, through 5 in Fig. 9.

For purposes of comparison the same effect may be obtained with electromechanical control as follows:

Referring to Fig. 6, assume the stopping purge contact closes to pick up 83P-1. Timer 2P-1 starts and times out; this in turn starts timer 2X. When 2X times out, 2X-1 picks up operating 2Z-1 in counting chain and starting timer Also 2X-1 resets timers 2P-1 and When 2Y times out, 2Y-1 picks up drop out 2X-1. This causes 2Z-2 the counting chain to operate and starts timer 2P-1 again. When 2P-1 ti out it drops out 2Y-1 and starts ti 2X. This cycle continues until all lays in the counting chain have b picked up at which time an outpu secured through the contacts of 83P-1 2Z-6. The counting chain is reset w 83P-1 drops out.

Fig. 7. Logic elements required for co of one turbine

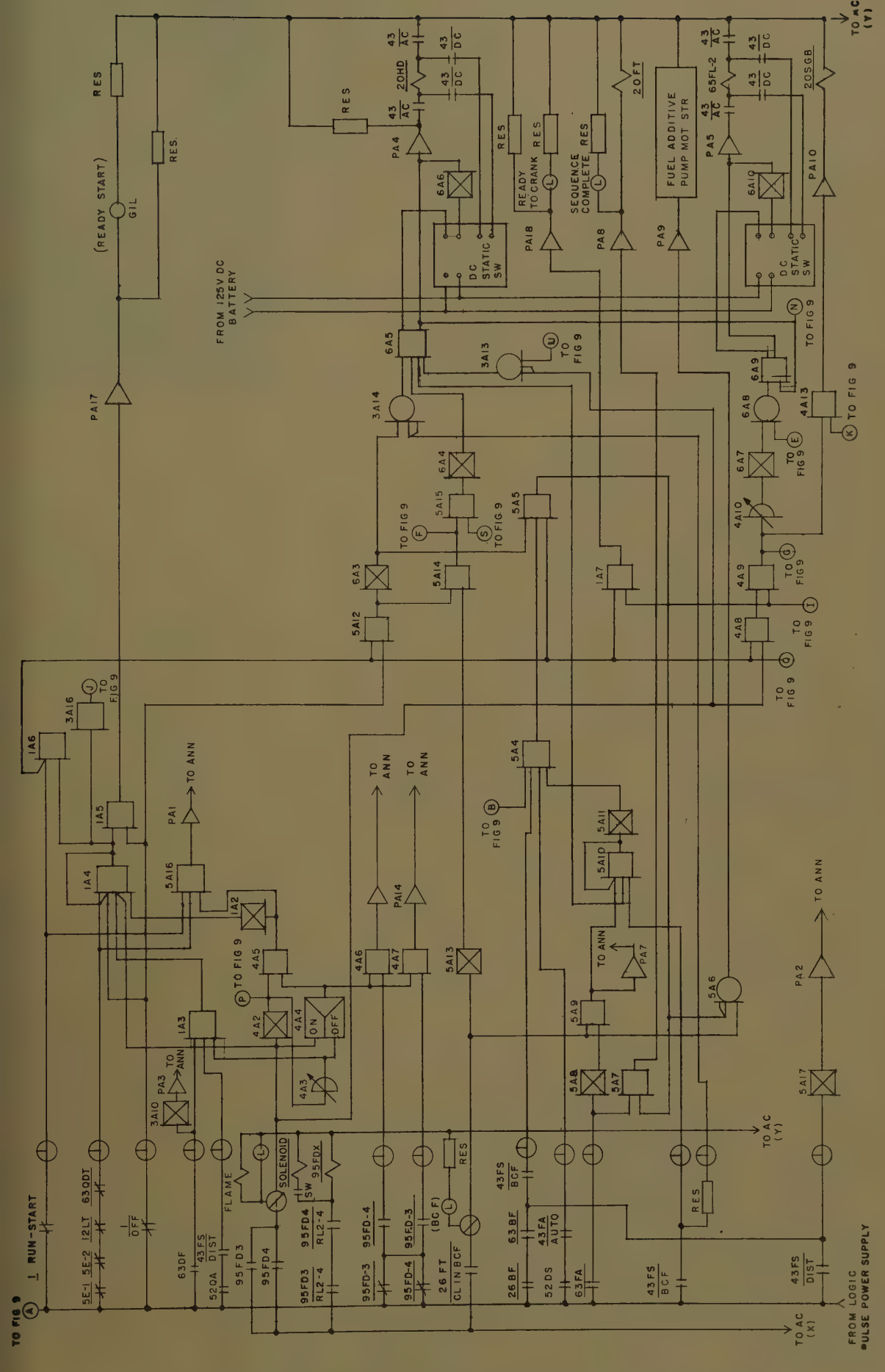


Fig. 8. Static control logic diagram (Part I)

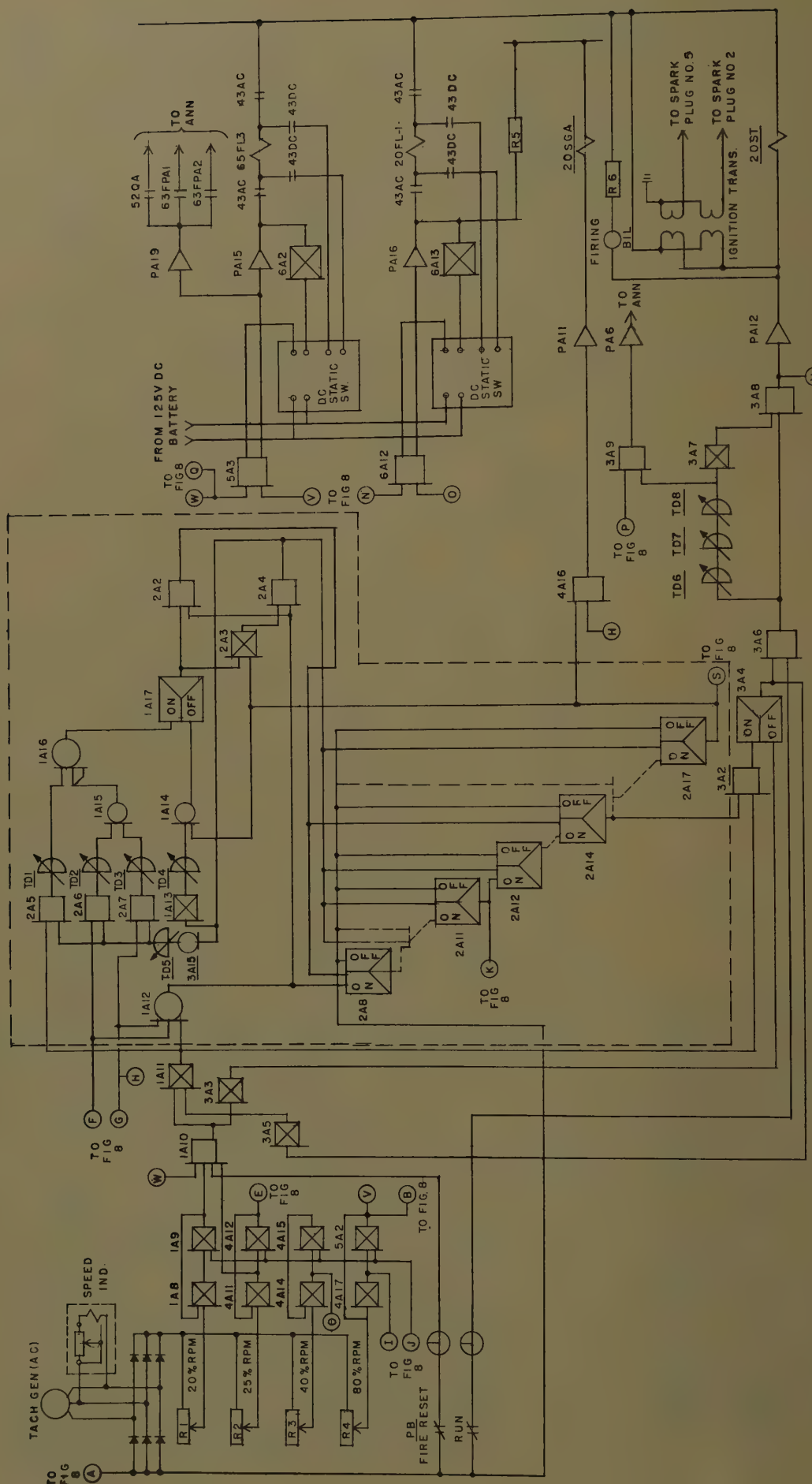
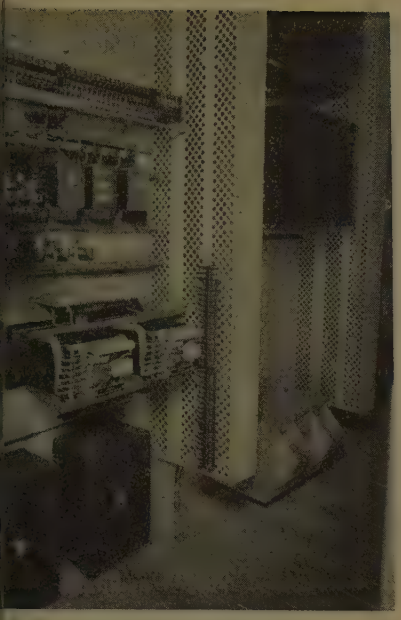


Fig. 9. Static control logic diagram (Part II)



10. Panel on which are mounted two regulators inside of duplex switchboard

evices 83A and 83P-2 operate the its in a manner similar to that de- ed above. The timers 2P-1, 2A, 2P-2, which each have maximum time y characteristics in the order of the ic time delay elements TD1 to TD5 ide a means of obtaining independ- y adjustable timing periods for each ne outputs.

neral Description of Operation

ig. 7 shows a close-up of the logic nents required for the control of one ne turbines. These units are mounted de the duplex control board at the of the left hand panel, as shown in part in Fig. 5. The time delay elements, among which are TD1 to TD5 whose operation has been described previously, are shown at the bottom of Fig. 7. The backs of two other panels in the duplex control board on which are mounted static switching devices are shown in Fig. 10.

Figs. 8 and 9 are the two parts of the logic control diagram for the control of one turbine. The operation of the individual circuits will not be explained in detail as has been done for the common timing circuit and counting chain shown in Fig. 9. The flow of signals between physically separated points on the diagram are indicated by terminating lettered circles. When the points appear on the two parts of the diagram these circles are so labeled. A general description of the operation of the control is as follows:

Starting

When all conditions are satisfactory for starting, the green lamp (GIL) located above the master control switch 1 in Fig. 8, will light. The operator turns the master control switch 1 to the START or RUN position and then engages the starting clutch. The cranking steam turbine brings the machine up to 20% speed, at which it runs for 3 minutes to purge the combustion system of combustible gases. If the master control switch was turned to START, the turbine will continue to crank at 20% speed and it will be necessary for the operator to turn the switch to the RUN position for the sequence to continue as outlined below.

If the master control switch is turned from OFF directly to RUN after the 3-minute purge period, a firing-timing circuit is energized, causing a "ready to fire" blue indicating lamp (BIL) in Fig. 9 to be lighted. At the same time the hydraulic dump valve is closed, the spark plugs are injected, and the ignition transformers and fuel limit solenoids are energized. After 45 seconds, the blue lamp is extinguished, the spark plugs are retracted, and the ignition transformers de-energized. If flame is not established within the 45-second firing period the turbine will continue to crank at 20% speed but will not proceed with the sequence. The annunciator will operate to indicate FAILURE TO FIRE and the hydraulic dump valve and fuel limit solenoids will be de-energized. Another attempt to fire may be made by the operator by depressing and releasing the FIRE RESET push button. This resets the controls for another purge and firing cycle.

If, during the 45-second firing period, flame is sensed by either flame detector, the hydraulic dump valve will remain closed and an amber indicating lamp is lighted, signifying flame has been established. Five seconds after flame is established, the fuel flow is reduced for a 2-minute warm-up period. At the end of this period the cranking turbine speed changing solenoid is energized, causing the cranking turbine to accelerate the unit. At 25% speed additional fuel is added. The turbine continues to accelerate for an additional 3-minute period. When the unit reaches 40% speed, additional fuel is added to bring the unit up to 80% speed. At this point maximum fuel is supplied to the unit causing

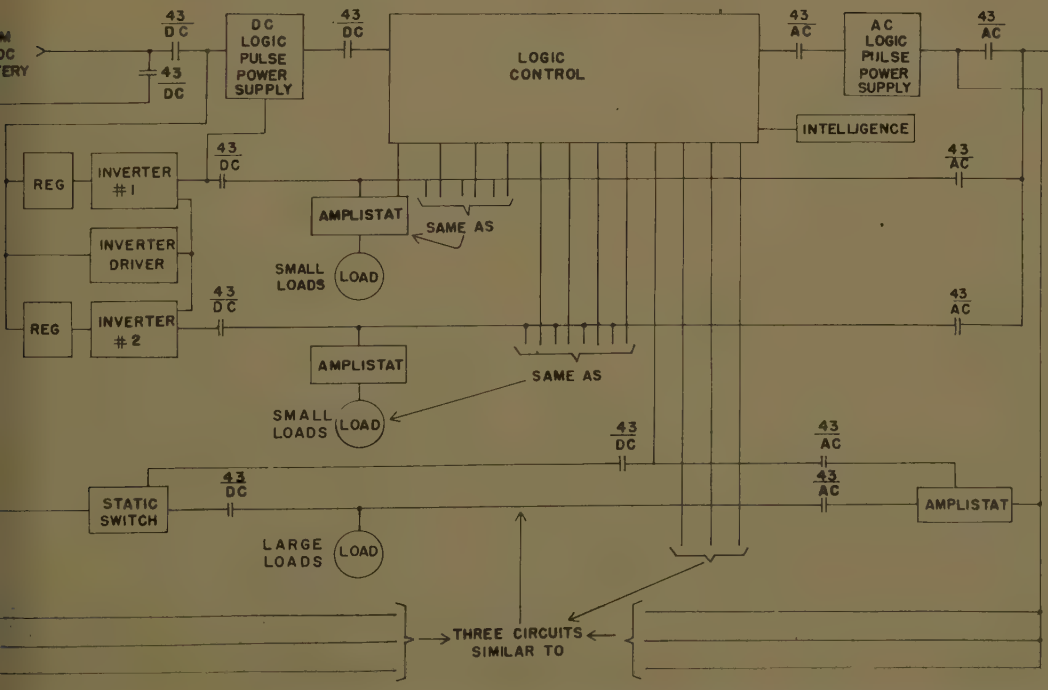


Fig. 11. Block diagram of static control for gas turbine

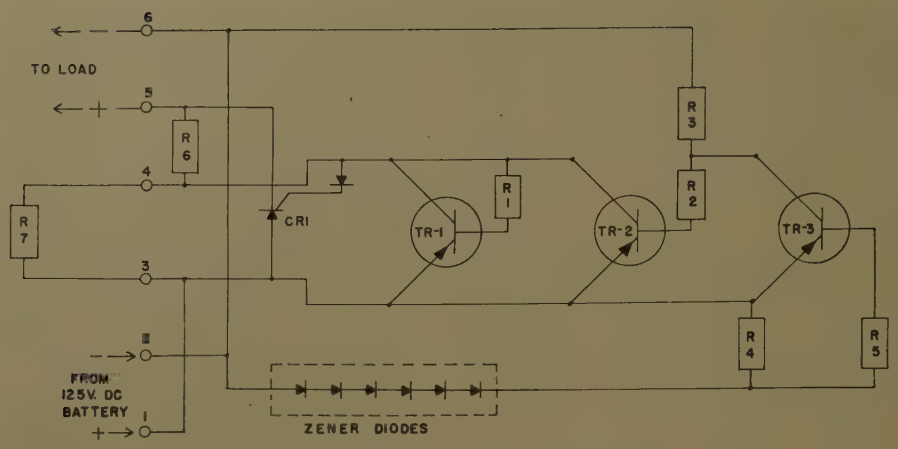


Fig. 12. Schematic diagram of static regulator

it to accelerate until the fuel flow is cut back by the governor. At some point between 40% and 80% speed the torque from the gas turbine will exceed the torque of the cranking turbine causing the two to become unclutched. The cranking turbine may then be shut down manually. When the machine is operating under control of the governor the speed may be changed by operating the speed control switch on the turbine control panel.

At any point after 80% speed is reached the unit may be transferred to residual fuel by turning the fuel selector switch to the appropriate position. The transfer will take place automatically provided conditions are satisfactory for operating on the residual fuel; otherwise, the machine will continue to operate on distillate fuel. Further, if, while the machines are operating on residual fuel, conditions should become unsatisfactory for continued operation with this fuel, the controls will automatically transfer back to distillate fuel.

After the turbines have reached operating speed they may be loaded as desired.

Shutdown Procedure

The operator unloads the turbine. He then turns the master control switch to the OFF position. The machine, if not already operating on distillate fuel, will automatically transfer to this type of fuel and will continue to operate for a 5-minute purge period. Upon completion of the purge cycle, the hydraulic dump valve is automatically opened; this cuts off the flow of fuel and causes the machine to stop.

Some Special Problems Encountered and New Devices and Equipment Developed

The static control and power amplifiers which were currently available, were designed for operation from an a-c control power source. Since the controls must also operate from a 125-volt battery source, it was necessary to develop several new static components.

The most straightforward solution for operating the controls from a battery would be to provide the necessary alternating current from the battery by means of a static-type inverter. However, upon closer analysis, it was found that the inverter, in order to handle the peak load requirements, would have to have considerable excess capacity. It was therefore decided to operate only the loads with relatively small peak demands from static inverters, the larger loads being operated directly from the battery through static switches.

The arrangement as actually used is shown in Fig. 11. As noted on the diagram two static-type series regulators were necessary for each unit to limit the d-c operating voltage automatically to 130 volts when charging the battery. Also, it is seen that the small loads are operated by the logic control through amplistats with inverters providing the necessary alternating current when operating from the battery. The loads with large peak demands are arranged to operate amplistats only when operating from

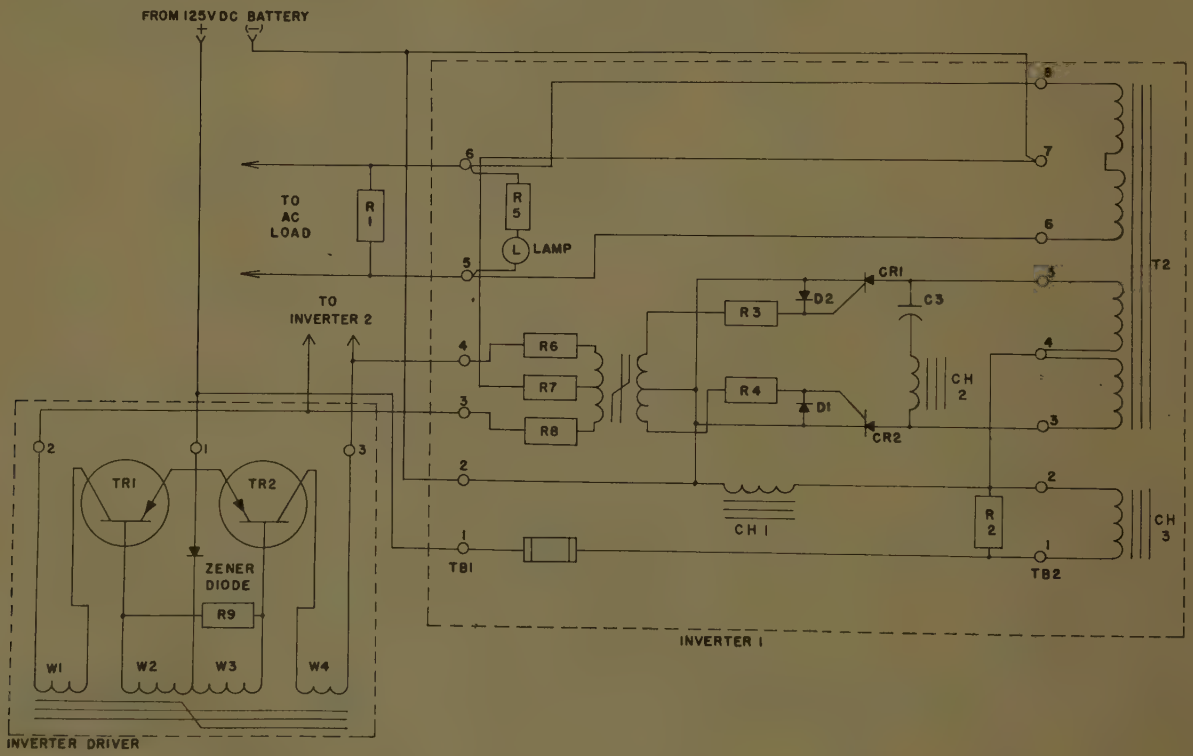


Fig. 13. Schematic diagram of static inverter and driver

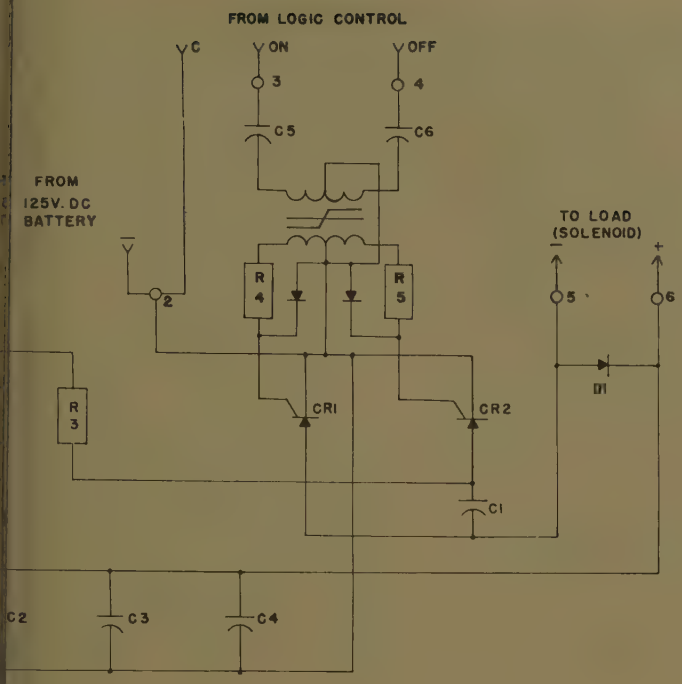


Fig. 14. Schematic diagram of static switch

control source. The static switches vide the necessary control when operng from the battery. A special transor–saturable-reactor combination switch vides the necessary pulse power to rate the logic control from the battery. Fig. 12 shows the schematic diagram one of the regulators. When the voltbetween terminals 1 and 2 exceeds the akover voltage of the ON Zener diodes, nsistor TR-3 is biased; this in turn ses the power transistors TR-1 and TR-2 off. The impedance between terminals 3 and 4 is thus increased to cause the output voltage at terminals 5 and 6 to be reduced. Instantaneous overcurrent protection for the power transistors is provided by the shunt R6 and the controlled rectifier CR1. When the drop across R6 becomes excessive, indicating an overcurrent, the controlled rectifier starts conducting to by-pass the current from the transistors. The rectifier continues to conduct until the fault protection in the form of fuses interrupts the circuit. The two regulators for one of the units are shown in the second row of devices from the bottom of the left-hand panel of Fig. 10.

Fig. 13 shows the schematic diagram covering a static inverter and its driver. The inverters are each rated to have an output of 500 voltamperes at 120 volts, 60 cycles with a 125-volt d-c input. The inverters use controlled rectifiers to switch opposite halves of the primary winding of transformer T2 across the battery alternately to produce an approximated square wave of alternating current at the secondary winding terminals. The transistor type of driver provides the necessary pulses to cause the controlled rectifiers to conduct in the proper sequence. It also provides for operation at a constant frequency over a wide range of input voltages. This is accomplished by using a Zener diode to control the time required for its transformer core to go through the excursions from positive to negative to positive saturation.

The schematic diagram used for one of the d-c static switches is shown in Fig. 14. Here again controlled rectifiers are used as the switching elements. They are designed to operate a-c type of solenoids which have inrush currents up to a maximum of 13 amperes and a minimum holding current of 0.6 ampere. The operation is as follows:

Assuming no control signals are applied

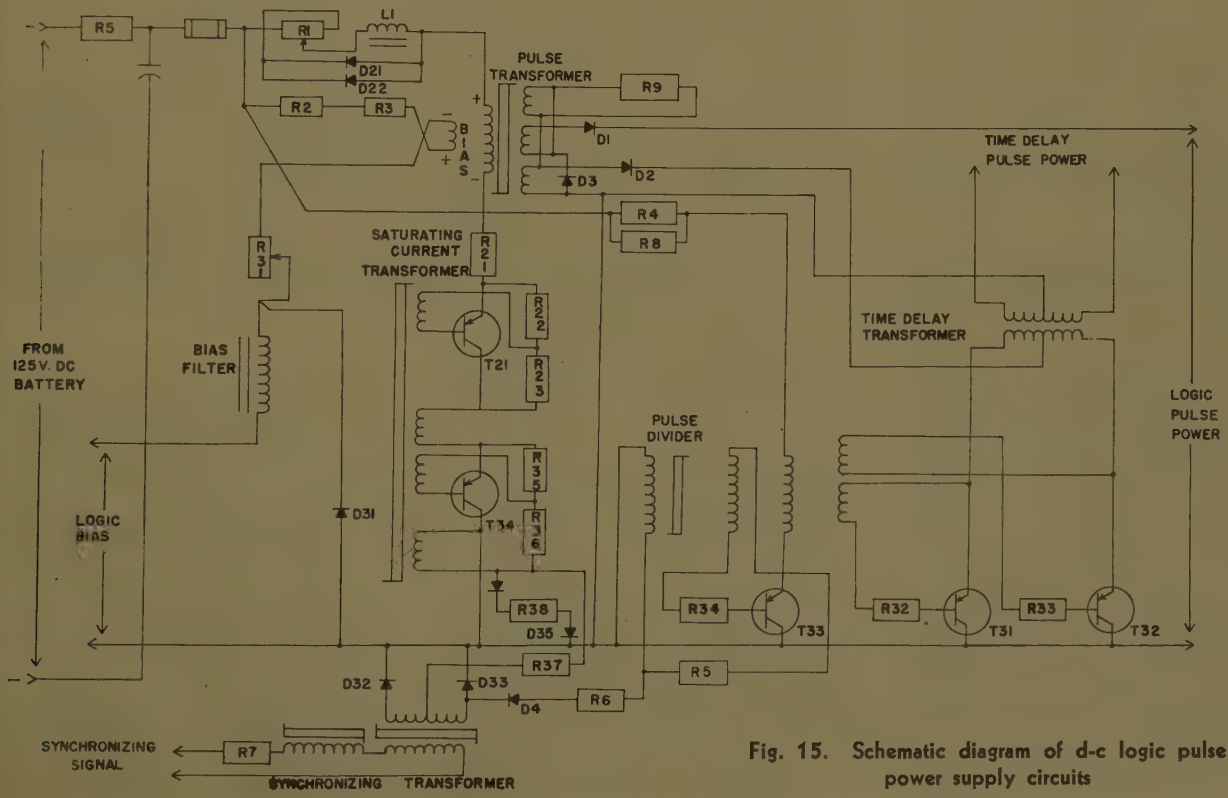


Fig. 15. Schematic diagram of d-c logic pulse power supply circuits

to terminals 3 and 4, the capacitors C2, C3, and C4 will charge up to the full source potential through resistors R1 and R2. By applying a positive signal between terminals 2 and 3, controlled rectifier CR1 fires. This permits capacitors C2, C3, and C4 to discharge their stored energy into the solenoid coil, causing it to pick up. The resistors R1 and R2 then pass enough steady-state current to hold in the solenoid after it is picked up. Capacitor C1 is charged to source potential through resistor R3 and controlled rectifier CR1. The solenoid is dropped out by removing the signal between terminals 2 and 3 and applying a positive signal between terminals 2 and 4. Controlled rectifier CR2 fires and momentarily puts inverse voltage from capacitor C1 on controlled rectifier CR1, thereby bucking its current to zero and giving it time to regain its nonconducting condition. As switching action takes place rectifier D1 conducts the solenoid current to prevent the self-induced voltage across the solenoid terminals from rising more than a few volts.

The d-c logic pulse power supply circuits are shown on Fig. 15. This device produces pulses of the proper magnitude, width, and polarity, and in the proper sequence to operate the static-type logic elements. The synchronizing signals are obtained from the output of one of the static inverters.

Referring again to Fig. 5, the static switches, the inverters and driver, and the d-c logic pulse power supply for one turbine are mounted on the left-hand panel in the above order, starting from the bottom of the panel. The amplistats for the unit are mounted on the adjoining panel to the right.

Placing of Equipment into Operation

Placing the control equipment into operation was not too different from the conventional electromagnetic type of controls.

Special training classes were conducted to familiarize operating and maintenance personnel with the static switching equipment. The static controls were assembled and tested at the factory. Therefore, it was only necessary to make field connections and provide a source of a-c and d-c power supply at the plant site. After all connections were completed, a thorough check was made for grounds and continuity as well as a visual check before power was applied to the system. The 115-volt a-c power source was first applied and a systematic operational check was made for each phase of the control sequence. Operating conditions were simulated by manually closing field contacts or applying an electric impulse at appropriate points. The electric impulse was initiated with dry cell batteries.

The sequence of operation was easy to follow by observing the indicating lamps on each static control element. The lamp is on when its respective element has an output. These lamps were very helpful in locating sources of trouble when an element failed to function. Another valuable aid in trouble shooting was a test box for checking the control and output currents of any specified element. The procedure for checking an element was to remove the element from its operating position and plug it into the test box. The test box was then plugged into the position from which the element was removed. The input (control) or the output current could be read on the milliammeter in the test box depending on the selection made.

Some erratic operation of the elements was encountered. For example, one of the Release Memory units in the counting chain sometimes would not return to the OFF condition after completion of the purge cycle before firing. The erratic operation was apparently eliminated but the cause was not definitely established because of the many corrective actions which were taken in the efforts to eliminate the erratic operation.

After the system was checked on a-c, it was switched to d-c and the operational check repeated. At this writing the turbines have been operating $2^1/_2$ months. During the second method of operation, the following turbine shutdowns were reported:

Turbine No. 1

Turbine No. 1 was shut down twice during the month, once for planned maintenance and once due to a fault in the d-c control power system. The control system was switched to a-c control power and operation continued while trouble on the d-c system was being corrected.

Turbine No. 2

Turbine No. 2 was shut down three times during the month due to the following reasons:

1. To perform maintenance work on the gas turbine itself.
2. For some undetermined reason, the turbine shut down. After an unsuccessful attempt to locate trouble in the control system, the turbine was restarted and ran satisfactorily.
3. The turbine was shut down for maintenance on the driven machine.

Nomenclature

Function Number

1 = master control switch
2A, 2P = time delay relays
5E = emergency trip
12LT = turbine overspeed trip
20FL = fuel limit valve solenoid
20FT = fuel transfer solenoid
20HD = hydraulic dump valve solenoid
20SGA (and B) = cranking turbine spe changing solenoids
20SI = spark plug injection pilot valve
26BT = bunker *C* fuel temperature trans
26FT = fuel temperature bunker *C* ope tion
43 = control power selector switch, a-c to
43FS = fuel selector switch
52DS = circuit breaker for dispersator
52QA = circuit breaker for auxiliary lubric ing pump
63BF = bunker *C* pressure low
63DF = low distillate fuel supply pressure
63FA = loss of additive flow
63FDA = fuel pump failure (alarm)
63QDT = loss of distillate lubricating pre sure
65FL = fuel limit
83A, 83P = automatic transfer relays
95FD = flame detector relay

Abbreviation

ANN = annunciator
AUTO = automatic
BCF = bunker *C* fuel
BIL = blue indicating lamp
C = capacitor
CH = inductance
CL = closed
CR = controlled rectifier
D = diode
Dist = distillate fuel
ER = electrical reset
GIL = green indicating lamp
L = indicating lamp
M, MTR = motor
OPER = operating
PA = power amplifier
R, RES = resistor
REG = regulator
RESET = resetting
STR = starter
SW = switch
T = transformer
TB = terminal block
TDC = time delay closing
TDO = time delay opening
TR = transistor
XYZ = auxiliary relays

References

1. Control of Gas Turbines for Powe Generation, D. C. Hoffmann, A. G. Mellor, N. Starkey. *AIEE Transactions* vol. 69, pt. II, 196 pp. 1637–42.

2. Automatic Control of Gas Turbines fo Natural Gas Line Pumping, C. R. Ingemanson Arne Loft, H. J. Wilt. *Ibid.*, pt. II (*Applications an Industry*), vol. 72, Nov. 1953, pp. 295–300.

3. Methods of Starting Gas Turbine-Gener tor Units, W. B. Boyum, R. W. Ferguson, J. Partlow. *Ibid.*, pt. III-A (*Power Apparatus an Systems*), vol. 73, Apr. 1954, pp. 322–28.

4. Centralized Control of 60,000 HP Gas Turbine Plant, H. D. Gibson, M. D. Todd *Power*, New York, N. Y., Jan. 1956, pp. 80–82.

5. Automatic Station Control, Supervisory and Associated Telemetering Equipment

C37.2–1956, American Standards Association, New York, N. Y., Oct. 1956.

THE APPLICATION OF STATIC SWITCHING CON- W. W. Litfin. *General Electric Publication 1366*, Schenectady, N. Y., Mar. 1957.

STATIC CONTROL APPLICATION MANUAL. *General Electric Publication GET 2909*, Aug. 1958.

STATIC SWITCHING TODAY, J. J. Duffy, S. . *Electrical Manufacturing*, New York, N. Y., 1959.

Discussion

Fahrenthold (Brown and Root, Inc., ston, Tex.): A unique application of ic control has been described in this er. The authors have done a most mendable job in completely covering the rse of this application from basic design inal check out of the installed unit.

Much information has been presented I would impose upon the authors by ing for amplification of the following nts.

n the section on "Why Static Control?" authors point out that the static control tem is more expensive than any other e which might have been used. They not give any indication of what the per- tage difference of cost might be between available systems. Information on this ect of the application, if available, would most desirable. It was also pointed out this section that one of the principal fea- es which made the static control system irable was the encapsulation of the tch units. Could hermetically sealed ays have been used with equal results?

In the section on Operating Requirements authors described the functions which re to be performed by the operators and se which were to be done automatically the control system. Under the manual on operator functions is listed, "Initiate transfer from distillate to residual fuel and back to distillate." However, the control scheme is to shut down the turbines automatically should any of the five abnormal conditions listed occur. For shutdown, the turbines must be switched from residual fuel to distillate. It is not clear whether, under the automatic shutdown, this transfer from one fuel to the other is done by the control system automatically or by the operator after he has been alerted by an alarm.

Should the control system lose its source of a-c supply, how does this affect its operation, since it also has a d-c supply and inverters? Would the units continue to function and if so for what period of time? This would be an important point in the original evaluation of the type drive to be used.

Would the economic evaluation be materially altered if the units were to operate for extended periods on distillate fuel? It would seem that this might be likely to occur.

I have assumed the combustion turbine and control system were a package unit. In this case the manufacturer is aware of the requirements of the control system. Perhaps the authors can advise whether we have reached the stage of the art where the logic units may be obtained separately and a specialized control be assembled by a user.

P. T. Carmack, E. M. Smith: The authors wish to thank Mr. Fahrenthold for his discussion, in which he has raised several pertinent questions.

It is difficult to arrive at a percentage difference in the cost between static and electromechanical types of control which would be generally applicable. Usually, a large percentage of the cost of static control is represented by the power amplifiers or output devices. Costs, therefore, are pretty much dependent upon the number and rating of the devices which would be required to accomplish a given task. In the past, cost of static control systems have run from 150 to 400% more than electromechanical devices required to perform similar functions. New developments and manufacturing techniques are, however, continually narrowing this cost differential.

It is true that hermetically sealed relays have been used in the past for controls located in a dusty and corrosive atmosphere. They are, however, available only in limited ratings which are not sufficient for controlling the larger loads in this system.

For any of the five listed abnormal conditions, it is recommended that the turbines be shut down immediately. Should such an operation occur while using residual fuel the control does not automatically transfer to distillate, since such a transfer, to be of benefit, would require continued operation of the machine. The control under any of these conditions operates the annunciator and shuts the machine down by cutting off its fuel supply.

Loss of the a-c control power source will cause a machine to shut down only if its controls are operating from this source at the time the loss occurs. When operating from the d-c control power source, loss of the a-c supply does not affect the operation of the machine. Further, there is provided a small steam-driven a-c stand-by unit which is of sufficient capacity to operate the battery charger. The battery is thus kept fully charged and operable during prolonged a-c outages.

The logic units used in this control system are a standard product which may be purchased separately. The new static components which were developed for this installation are not generally available.

Nonlinear Compensator for a Piecewise Linear Second-Order Feedback System

MICHAEL ATHANASSIADES
STUDENT MEMBER AIEE

OTTO J. M. SMITH
MEMBER AIEE

nopsis: The nonlinear computer for a cond-order system preceded by a saturat- amplifier, with significant linear zone, s designed through the transformation of e zero-force curve and the geometry of the ase plane into the topology of the non- ear computer. A finite-gain system was us designed for the requirement of mini- im settling time for step input functions. ie method of design was based upon the vision of the phase plane into three re- ons of operation: a saturated one, and o linear ones. The equation of the zero- ce curve was that of a parabola. The nlinear compensator had as its output the n of the error signal and a nonlinear func- n of the error rate. Different systems re compared in order to determine the fastest one. One of them proved to be superior to all the others for a spectrum of different input step magnitudes.

THE CONTROL of the double-integral system with nonlinear compensation was examined by McDonald[1] and Hopkin[2] for the case of a very-high-gain amplifier or of a relay used for the power amplification necessary. The purpose of this paper is to modify the nonlinear computer used by McDonald in order to make the system respond satisfactorily when the power amplifier has a finite gain and a significant linear zone. The block diagram for the relay double-integral system is shown in Fig. 1. All values are normalized. The purpose of this research was to determine the topology of the nonlinear computer shown in Fig. 1. One form of nonlinear computer is shown in the diagram of Fig. 2.

The switching curve for the system of Fig. 1 is given by the expression

$$(\text{sgn } \dot{e})\dot{e}^2+2e=0 \qquad (1)$$

If the nonlinear computer of Fig. 2 is identical to that of Fig. 1, then equation 1 is the locus of all points in the phase for which $F = 0$. Let m be the output of the nonlinear computer and the input to the saturating amplifier. It is also called

Paper 60-116, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 26, 1959; made available for printing December 16, 1959.

MICHAEL ATHANASSIADES is a student and OTTO J. M. SMITH a professor at the University of California, Berkeley, Calif.

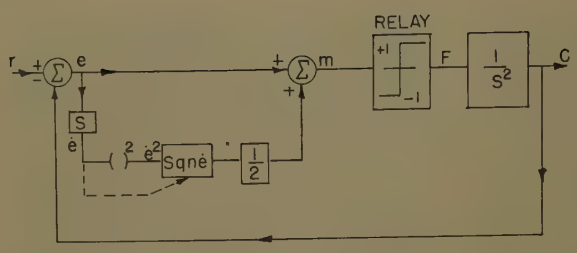


Fig. 1 (left). Block diagram for the relay "bang-bang" servo-mechanism

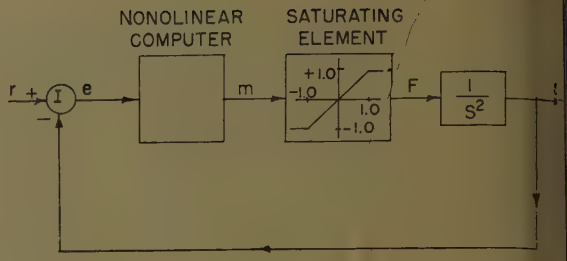


Fig. 2 (right). Suggested topology for the system

the "manipulated variable." The locus of all points in the phase plane for which $F = m = 0$ will be named the "zero-force curve" (ZFC). The locus of all points in the phase plane for which $m=1$ will be called the "+1 linear boundary," and the locus of all points for which $m=-1$ is called the "−1 linear boundary." The two linear boundaries divide the phase plane into two regions of operation: region *1*, the saturated region, and region *2*, the linear region. These two regions are shown in Fig. 3.

For a step input function R_o the trajectory equation in region *1* is given by

$$(\text{sgn } \dot{e})\dot{e}^2 - 2(e - R_o) = 0 \qquad (2)$$

the same as for the relay system of Fig. 1.

The differential equation of the system in region *2* is

$$\dot{e}\frac{d\dot{e}}{de} + \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 + e = 0 \qquad (3)$$

Equation 3 is a form of Bernoulli's equation and may be solved quite easily. The solution is

$$\dot{e}^2 = Ke^{-(\text{sgn } \dot{e})\, e} - 2(\text{sgn } \dot{e})e + 2 \qquad (4)$$

where the constant K depends on the initial conditions or on the conditions at the end of the trajectory of region *1*. Fig. 3 also shows some typical trajectories for small step inputs. It is seen that the response is not good since high overshoot and long settling time are inevitable. For large step inputs, however, the system response is satisfactory.

Modification of the Phase-Plane Design

In order to compensate for the oscillatory response of the system the nonlinear computer of Fig. 2 was modified. This modification is based on the introduction of a second linear zone of operation about the origin of the phase plane. This modification in the geometry of the phase plane was accomplished in two steps.

Step 1: The displacement of the zero-force curve along the error axis.

Let the ZFC, given by equation 1, and the linear boundaries be displaced from the origin by an amount $1+a$ along the error axis of the phase plane to the right in the lower half-plane and to the left in the upper half-plane of the origin as shown in Fig. 4. Thus the ZFC has the equation

$$e + \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 + (\text{sgn } \dot{e})(1+a) = 0 \qquad (5)$$

If a physical system were built having the phase-plane geometry of Fig. 4 it would have a steady-state error of position of magnitude $1+a$ for any step input. Clearly, this is very undesirable. The ZFC must be modified so that it passes through the origin of the phase plane, in order to eliminate any steady-state errors.

Step 2: The introduction of a second region of operation.

Let the ZFC of the system be described by equation 5 as long as the magnitude of error rate is larger than an amount $\dot{e}_m$.

Let the ZFC be described by the equation

$$\dot{e} = f(\dot{e}) \qquad (6)$$

in the region $-\dot{e}_m \leqslant \dot{e} \leqslant +\dot{e}_m$, where $f(e)$ is any linear or nonlinear function of the error satisfying the following conditions:

1. The $f(\dot{e})=0$, when $\dot{e}=0$. This satisfies the requirement of zero steady-state error.
2. The function $f(\dot{e})$ must coincide with the ZFC of region *2* at the points $\dot{e}=\pm\dot{e}_m$. Thus, no discontinuities are allowed in the ZFC.
3. The function $f(\dot{e})$ must be monotonic and continuous. Slope discontinuities are allowed for a finite number of points.

The resultant phase plane, with the three regions of operation, is shown in Fig. 5.

The equations of the linear boundaries in region *3* are

$$+1 \text{ linear boundary: } e = f(\dot{e}) + 1 \qquad (7$$

$$-1 \text{ linear boundary: } e = f(\dot{e}) - 1 \qquad (8$$

Thus, the nonlinear computer may be designed from Fig. 5.

In region *2*, the quasilinear region of operation due to the linear band of the amplifier, the manipulated variable is

$$m = e + \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 + (\text{sgn } \dot{e})(1+a) \qquad (9$$

In region *3*, the new quasilinear region of operation due to the linear band about the origin, let the manipulated variable be

$$m = e + \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 + \phi(\dot{e}) \qquad (10$$

where the function $\phi(\dot{e})$ is to be determined.

When $m=0$, $e=f(\dot{e})$. Substituting into equation 10

$$0 = f(\dot{e}) + \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 + \phi(\dot{e}) \qquad (11$$

therefore,

$$\phi(\dot{e}) = -f(\dot{e}) - \frac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 \qquad (12$$

Fig. 6 shows the block diagram of the nonlinear computer required. It is evident that the only difference between Fig. 6 and Fig. 1 is the addition of a single nonlinearity ϕ in the error rate channel.

The output of the nonlinearity ϕ is the function $\phi(\dot{e})$ defined as follows from equations 9 and 12:

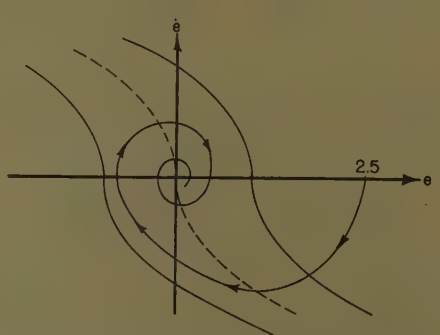


Fig 3. Phase-plane trajectory for the uncompensated system due to a step input

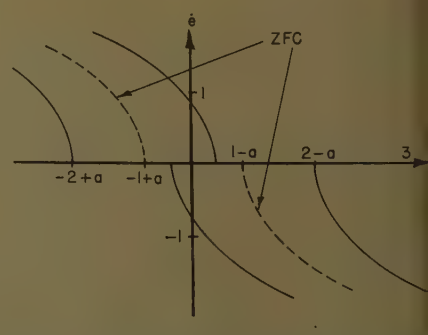


Fig. 4. Symmetrical phase plane due to the shifting of the ZFC by ±(1+a). The value of "a" is negative

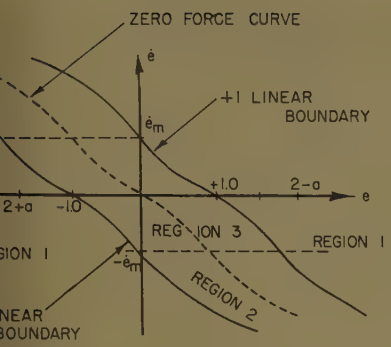


Fig. 5 (left). Phase-plane portrait illustrating the three regions of operation

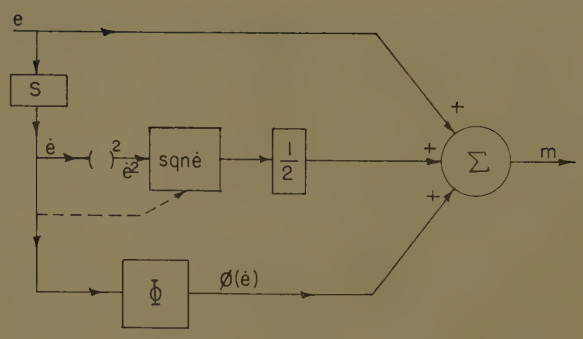


Fig. 6 (right). Topology of the nonlinear computer for the general system

$$\begin{cases} (\operatorname{sgn}\dot{e})(1+a) & \text{for } |\dot{e}| > \dot{e}_m \quad (13A) \\ -f(\dot{e}) - \frac{1}{2}(\operatorname{sgn}\dot{e})\dot{e}^2 & \text{for } |\dot{e}| \le \dot{e}_m \quad (13B) \end{cases}$$

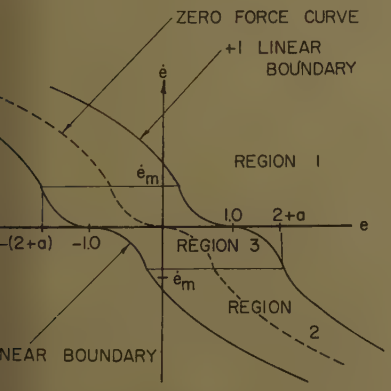
e $f(e)$ is the equation of the ZFC in
n *3*.
ie differential equation of the system
e-plane trajectory in region *3* is

$$+f(\dot{e})=0 \qquad (14)$$

ie necessity of introducing a second
r zone of operation was to reduce the
unt of overshoot and the settling
. From the general method of de-
presented it follows that if the amount
is decreased or if the value of the
tity $1+a$ is increased the system will
me less oscillatory. Of course, its
onse depends largely on the shape of
ZFC.
ius, it has been demonstrated that
a general case the geometrical con-
ation of the phase plane may be
sformed into the required topology of
nonlinear computer through the use
e ZFC.

ie Applications of the General heory

this section an equation will be as-
ed for the zero-force curve of the
l region of operation. The equations



7. Phase-plane portrait for a particular application

of the phase-plane curves will be developed, and, finally, some experimental results will be presented, as obtained through the use of an analog computer.

The form of the ZFC, as well as that of the linear boundaries, is shown in Fig. 7. From that figure the division of the phase plane into its three regions is clear. It is seen that the ZFC is composed of four segments: two in the upper phase plane and two in the lower phase plane. The ZFC is continuous in the phase plane, although at the boundaries of regions *2* and *3* there is a discontinuity in the slope of the ZFC. The ZFC, moreover, passes through the origin of the phase plane. The properties of the ZFC are

1. The ZFC, in the entire phase plane, is a set of four parabolas, two with vertices at the points $(\pm\dot{e}_m, \pm e_z)$, passing through the origin of the phase plane and lying in region *3*. The other two parabolas lie in region *2* and they have their vertices at the points $(\pm(1+a), 0)$.

2. The portions of the parabolas of region *3* which would extend into regions *2* and *1* of the phase plane are eliminated, and the parabolas are terminated at the point $\pm\dot{e}_m$.

3. The portions of the parabolas of region *2* which would extend into regions *3* and *1* are eliminated.

4. The two parabolas of region *3* are tangent to each other at the origin of the phase plane, and they have infinite slope at the point $\dot{e}_m$.

5. The entire set of the four parabolas is continuous at every point.

For any value of $\bar{a}$ and $\dot{e}_m$ the equation of this particular ZFC is

$$e = (\operatorname{sgn}\dot{e})\dot{e}^2 \frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m^2} - \frac{\dot{e}_m^2+2(1+a)}{\dot{e}_m}\dot{e} \qquad (15)$$

The derivation of equation 15, based upon the properties of ZFC, is shown in Appendix I.

Once the equation of the region *3* ZFC is known, the nonlinear computer may be designed immediately.

Substituting equation 15 into equations 13(A) and 13(B), the required function $\phi(\dot{e})$ is obtained, and, after some manipulation, is

$$\phi(\dot{e}) = \begin{cases} (\operatorname{sgn}\dot{e})(1+a) & \text{for } |\dot{e}| > \dot{e}_m \quad (15A) \\ -(\operatorname{sgn}\dot{e})\dot{e}^2 \frac{\dot{e}_m^2+(1+a)}{\dot{e}_m^2} + \frac{\dot{e}_m^2+2(1+a)}{\dot{e}_m}\dot{e} & \text{for } |\dot{e}| \le \dot{e}_m \quad (15B) \end{cases}$$

The block diagram of the nonlinear computer for this system is shown in Fig. 8. There are three channels of nonlinear operation in the error-rate channel: *A*, *B*, and *C*. Channel *A* fixes the large signal

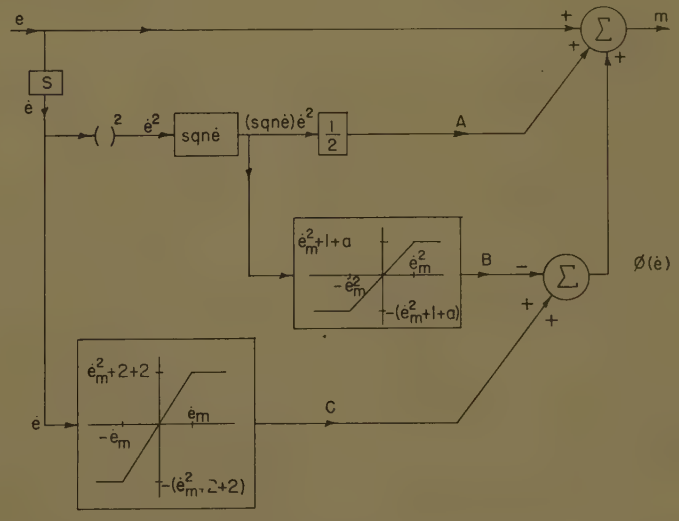


Fig. 8. The nonlinear computer designed from equations 15(A) and 15(B)

Table I. Value of the "Goodness" Parameter H as a Function of $\bar{a}$ and $\dot{e}_m$

$\bar{a}$	$\dot{e}_m$	H
−0.4	+0.4	3.24
−0.4	+0.5	2.60
−0.4	+0.7	2.29
−0.4	+1.0	3.30
−0.3	+0.4	3.27
−0.3	+0.5	3.90
−0.3	+0.7	2.52
−0.3	+1.0	3.39
−0.2	+0.4	3.06
−0.2	+0.5	1.70*
−0.2	+0.7	2.50
−0.2	+1.0	3.39
−0.1	+0.4	3.39
−0.1	+0.5	3.88
−0.1	+0.7	3.30
−0.1	+1.0	3.36
0	+0.4	4.47
0	+0.5	3.66
0	+0.7	3.53
0	+1.0	3.18

* Minimum.

response of the system. Channel B mainly controls the system behavior about the point $\dot{e}_m$. Channel C controls the very small signal behavior of the system. This is evident since the output signals of channels A and B are some functions of the square of the error rate, while the output signal of channel C is some function of the error rate. Near the origin the square of the error rate is negligible compared with the error rate, hence channel C is the most important one in that region.

If all the nonlinearities are constructed through the use of biased diode networks, the three nonlinearities in the error-rate channel may be combined into one only for the actual system.

The nonlinearities required are of the simple limiter type and the maximum signal levels, as well as the gains of their linear zones, depend on the values of the quantities $\bar{a}$ and $\dot{e}_m$.

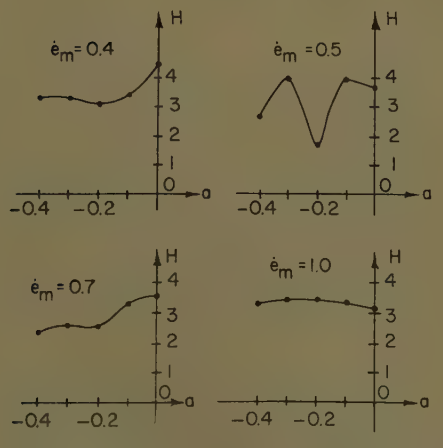


Fig. 9. The dependence of the goodness factor H upon the constants "a" and $\dot{e}_m$

Experimental Results

The system was tested through the use of an analog computer. The input signals were step functions of magnitude of 0.5, 1.0, 1.5, 2.0, and 2.5 units. The reason that this limited range of signals was considered was that for large step inputs the response of the system approaches that of the dual-mode relay servo, because in that case the width of the linear zone of the saturating element is small compared with the signal magnitude. The range of the quantities $\bar{a}$ and $\dot{e}_m$ selected for test were

$a = -0.4, -0.3, -0.2, -0.1$, and 0

$\dot{e}_m = 1.0, 0.7, 0.5$, and 0.4 units

The system was tested for each combination of the five values of R_o, the magnitude of the input step, and of $\bar{a}$ and $\dot{e}_m$. The error, error rate, output, and force F were recorded using a Sanborn recorder. From each curve the rise time, the peak overshoot ratio, and the settling time were obtained. The settling time was defined as the time required for the error, after an input step was applied, to diminish to 5% of the magnitude of the input step.

Since the system is gain limited, the response depends upon the magnitude of the input. It was found that, except for very small signals, the settling time increased as the magnitude of the input step increased. In order to select a criterion for the best design possible the following must be considered:

1. For a positioning system such as this the most important criterion of its transient response is its speed. Therefore, the best system is the one with the smallest settling time.

2. The absence of overshoot is desirable, but its importance is secondary only to that of the settling time. For example, a system with 2% overshoot and a 10-second settling time is inferior to a system with 10% overshoot but only 5-second settling time.

3. Since the system is gain limited the criterion of selection must compensate for the long settling time for large step inputs.

4. The probability of any step, between the values of 0.5 and 2.5 units, appearing at the input is assumed to be equal.

It is well known that the best system that can be built is the dual-mode relay system described in the beginning of this paper. This system has, theoretically, infinite gain and its step response is dead beat, i.e., finite settling time and zero overshoot. The settling time naturally is a function of the input R_o. The time required for the error of the dual-mode system to become zero is given by the expression

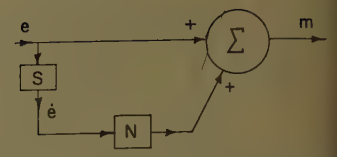


Fig. 10. Required topology of the nonlin computer

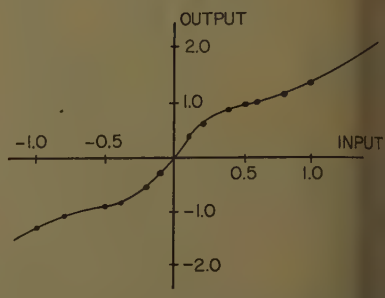


Fig. 11. Input-output characteristics of nonlinearity N in Fig. 10 for the best des

$$t_{dm} = 2(|R_o|)^{1/2} \quad (1$$

The derivation of equation 16 is given Appendix II.

In the system being studied, for ea combination of $\bar{a}$ and $\dot{e}_m$ there is a settli time $t_s(R_o)$ as a function of the inp Define the quantity

$$W_{R_o} = \frac{t_s(R_o)}{t_{dm}} \quad (1$$

The value of W is an indication of t speed of the system being studied co pared with the dual-mode one. If W 1.0, then the system is as good as the rel servo.

Also, define the quantity H:

$$H = \frac{1}{5}(W_{0.5} + W_{1.0} + W_{1.5} + W_{2.0} + W_{2.5}) \quad (1$$

The best design, i.e., the best combi tion of $\bar{a}$ and $\dot{e}_m$, is defined as the one w the minimum H.

Table I shows the value of H as a fu tion of $\bar{a}$ and $\dot{e}_m$. From the table it obvious that the minimum H is the c produced by the values $a = -0.2$ a $\dot{e}_m = 0.5$.

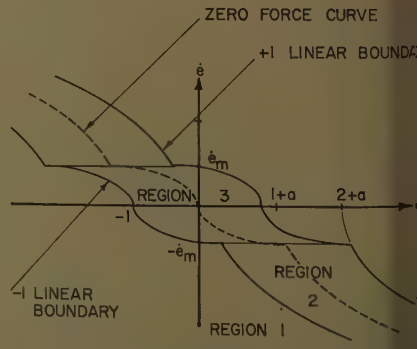


Fig. 12. Phase-plane portrait for another p sible application

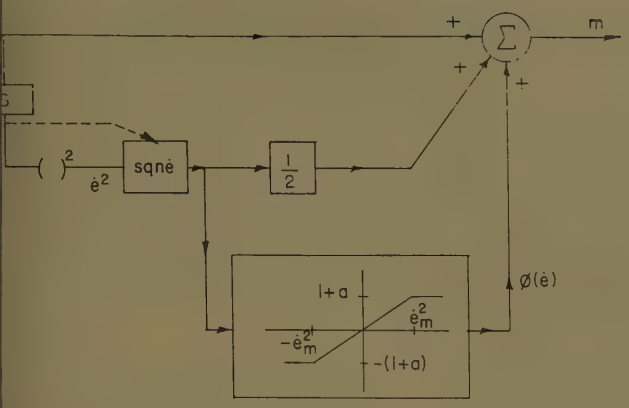


Fig. 13. Nonlinear computer designed from equations 20(A) and 20(B)

g. 9 shows the value of H for each cular design. One of the characteris- of this particular system is that for combination of $\bar{a}$ and $\dot{e}_m$ there is an imum" value of magnitude R_{op}, for h the step response is dead beat for ngineering purposes. This dead-beat nse has the minimum settling time zero overshoot. The inverse state- is not true, because of every value of here are one or more combinations of d $\dot{e}_m$ for which the step response is beat.

g. 10 shows the block diagram for best design with one nonlinearity N e error rate channel. Fig. 11 shows input-output characteristics of the inearity N. As it is seen, it is quite to construct it using biased diodes the gain required is finite and not ssive. Furthermore, deviations of the al components from their theoretical es are not critical as was the case with dual-mode relay servo.

e More Applications to Specific xamples

me other systems were examined in tion to the one just described. ir design is described here briefly.

TEM WITH ZFC A PARABOLA WITH ERTEX AT THE ORIGIN

or this system the zero-force curve of n 3 was selected to be a set of four parabolas, two with their vertices at the origin, lying in region 3, and terminating at the point $\pm\dot{e}_m$. The other two parabolas, lying in region 2, have their vertices at the points $(\pm(1+a), 0)$.

The parabolas of region 3 do not extend into regions 1 and 2, and the parabolas of region 2 do not extend into regions 1 and 3.

The phase-plane portrait for this system is shown in Fig. 12. It can be seen that the ZFC is continuous, that there is a slope discontinuity at $\pm\dot{e}_m$, and that the ZFC has infinite slope at the origin. Furthermore, the ZFC is concave to the right for the lower phase plane, and concave to the left for the upper phase plane.

The equation of the ZFC is, in region 3,

$$e = -\frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m^2}(\text{sgn } \dot{e})\dot{e}^2 \qquad (19)$$

The function $\phi(\dot{e})$ is obtained by insertion of equation 19 into equation 13:

$$\phi(\dot{e}) = \begin{cases} (\text{sgn } \dot{e})(1+a) & \text{for } |\dot{e}| > \dot{e}_m \quad (20A) \\ (\text{sgn } \dot{e})\dot{e}^2 \dfrac{1+a}{\dot{e}_m^2} & \text{for } |\dot{e}| \le \dot{e}_m \quad (20B) \end{cases}$$

The nonlinear computer is shown in Fig. 13 and the additional nonlinearity is of the limiter type, in the error-rate channel. From Fig. 12 it is seen that the ZFC has an infinite slope at the origin of the phase plane. Therefore, the system is conservative for very small inputs, and hence oscillatory. This system was also tested experimentally. Proceeding as in the previous section, it was found that its best design was the design with minimum H equal to 2.56. Hence this second system is inferior to the best design of the first system.

ZFC A STRAIGHT LINE IN REGION 3

If the ZFC is a straight line in region 3, the system is linear. Such a case is shown in Fig. 14.

The equation of the ZFC is

$$e = -\frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m}\dot{e} \qquad (21)$$

The required function $\phi(\dot{e})$ is

$$\phi(\dot{e}) = \begin{cases} (\text{sgn } \dot{e})(1+a) & \text{for } |\dot{e}| > \dot{e}_m \quad (22A) \\ -\dfrac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m}\dot{e} - \dfrac{1}{2}(\text{sgn } \dot{e})\dot{e}^2 & \text{for } |\dot{e}| \le \dot{e}_m \quad (22B) \end{cases}$$

The schematic for the nonlinear computer is shown in Fig. 15. Note that the added limiter in the error-rate channel subtracts the function $(1/2)(\text{sgn } \dot{e})\dot{e}^2$ when the system operates in region 3.

Conclusions

It has been shown that the addition of a third region of operation in the phase plane improves the transient response of the second-order system. Using a general case, the geometry of the phase plane has been transformed into the topology of the nonlinear computer through the use of of the zero-force curve. This concept may be used for the satisfactory design of any second-order system.

The authors plan to present another paper which demonstrates the additional usefulness of the ZFC by which a nonlinear system is described in terms of an equivalent linear one with moving poles and zeros in the s-plane. The authors believe that a transformation may be found

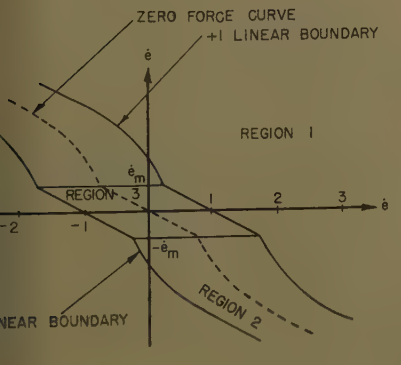


Fig. 14 (left). Phase-plane portrait for a system with a linear ZFC in region 3

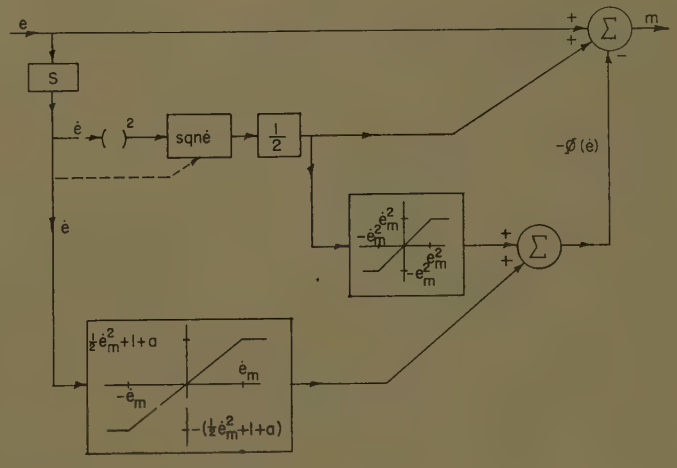


Fig. 15 (right). Nonlinear computer designed from equations 22(A) and 22(B)

that will link the transient method of design with the "describing function" method of frequency response used extensively in nonlinear systems of similar nature.

Appendix I

From Fig. 7:

$$e_z=\frac{1}{2}\dot{e}_m^2+1+a \tag{23}$$

For the lower half of the phase plane the zero-force curve has the form

$$(\dot{e}+\dot{e}_m)^2=Q(e-e_z)=Q(e-\frac{1}{2}\dot{e}_m^2-1-a) \tag{24}$$

$$Q=\frac{\dot{e}^2+2\dot{e}\dot{e}_m+\dot{e}_m^2}{e-\frac{1}{2}\dot{e}_m^2-1-a} \tag{25}$$

Since the ZFC passes through the origin of the phase plane

$$Q=-\frac{\dot{e}_m^2}{\frac{1}{2}\dot{e}_m^2+1+a} \tag{26}$$

Therefore the ZFC for the lower phase plane is given by

$$(\dot{e}+\dot{e}_m)^2=-\frac{\dot{e}_m^2}{\frac{1}{2}\dot{e}_m^2+1+a}\left(e-\frac{1}{2}\dot{e}_m^2-1-a\right) \tag{27}$$

or

$$e=-\dot{e}^2\frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m^2}-\frac{\dot{e}_m^2+2(1+a)}{\dot{e}_m}\dot{e} \tag{28}$$

Similarly, for the upper phase plane the ZFC is found to be

$$e=+\dot{e}^2\frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m^2}-\frac{\dot{e}_m^2+2(1+a)}{\dot{e}_m}\dot{e} \tag{29}$$

Combining equations 28 and 29 one obtains

$$e=(\operatorname{sgn}\dot{e})\dot{e}^2\frac{\dot{e}_m^2+2(1+a)}{2\dot{e}_m^2}-\frac{\dot{e}_m^2+2(1+a)}{\dot{e}_m}\dot{e} \tag{30}$$

Appendix II

Equation 16 is derived as follows:

$$t_{am}=\int dt=\int\frac{de}{dt}\,dt=\int de/\dot{e} \tag{31}$$

For a step input R_o the trajectory has the equation given by equation 2. When $e=R_o/2$ the relay switches and the final trajectory is given by equation 1. Therefore,

$$t_{am}=\left|\int_{R_o}^{R_o/2}\frac{de}{(2(\operatorname{sgn}\dot{e})(\dot{e}-R_o))^{1/2}}\right|+\left|\int_{R_o/2}^{0}\frac{de}{(-(\operatorname{sgn}\dot{e})2e)^{1/2}}\right|$$
$$=(|R_o|)^{1/2}+(|R_o|)^{1/2}$$
$$=2(|R_o|)^{1/2}$$

References

1. Nonlinear Compensator for a Piecewise Linear Second Order System, Michael Athanassiades. *MS Thesis*, University of California, Berkeley, Calif., June 1959.

2. Nonlinear Techniques for Improving Servo Performance, D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. pp. 400–21.

3. A Phase-Plane Approach to the Compensation of Saturating Servomechanisms, A. Hopkin. *AIEE Transactions*, vol. 70, pt. I, 19 pp. 631–39.

4. A Dual-Mode Servomechanism Utilizing Saturation Switching, H. R. Weed, F. Weimer. *Ibid.*, pt. II (*Applications and Industry*, vol. 77, 1958 (Jan. 1959 section), pp. 590–95.

5. Feedback Control Systems (book), O. J. Smith. McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

Discussion

G. J. Thaler (U. S. Naval Postgraduate School, Monterey, Calif.): The design of deliberately nonlinear control systems is, at present, more of an art than a science. This paper presents a scientific approach to the design of a specific class of deliberately nonlinear control systems, and appears to be an important step in a desired direction.

The basic concept of a saturating system with linear mode bounded by saturated deceleration trajectories was first proposed by McDonald,[1] extended to higher order systems by Neiswander and MacNeal,[2,3] and translation of the boundaries with respect to the origin of the phase plane is evident in the work of Rauch and Howe.[4] None of these authors are concerned with the design of the nonlinear compensator, however, and it is in this area that the present paper is a definite contribution.

The authors chose to consider the case of a double-integral system, which corresponds to a positioning servomechanism with no damping. In the practical case damping is present, the equations are more complicated, and the saturation boundaries have a different shape. When this is the case the system parameters must be determined accurately, and the design computations are more laborious. The specific results shown by the authors in Fig. 9 would not apply but the techniques leading to these results should still be valid.

With regard to Fig. 4 and equation 5 the authors state: "If a physical system were built having the phase-plane geometry of Fig. 4 it would have a steady-state error of position of magnitude $1+a$ for any step input." This does not seem to be correct. For any step of magnitude greater than $1+a$ the phase trajectory must cross the ZFC and hit the e-axis between the two ZFC's. When the state point reaches the e-axis the motion is at an "end point" and theoretically no phase trajectory can be extended from this location. Using physical interpretations based on normal topology for the control loop (which includes Figs. 6, 8, 10(A), 12, and 14) the state point indicates zero velocity and finite error. The error signal therefore establishes a signal, m, which activates the system and forces an equilibrium condition at the origin of the $\dot{e}$ versus e plane, that is, the steady-state error is zero. The action of the system during this motion might be compared to that of a relay servo with too much derivative control which results in relay chatter. In this case the motion must be much slower, however, and one must agree with the author's conclusion that the performance would be undesirable.

A number of interesting modifications are possible. Obviously tachometer feedback may be substituted for the error differentiator since they are identical for step inputs (except for an impulse and sign reversal, and the impulse is of no consequence here). It might also be

Table II. Transient Response of System with Nonlinear Computer of Fig. 8

$\dot{e}_m$	a	Step Input Magnitude	Rise Time	Overshoot, %	Settling Time	Comments
0.5	0.2	0.5	∞	0	5.95	
0.5	0.2	1.0	2.85	4.0	2.45	
0.5	0.2	1.5	2.90	4.9	3.80	This is the optimum design.
0.5	0.2	2.0	2.87	7.5	5.00	
0.5	0.2	2.5	2.70	13.0	6.00	
0.7	0.4	0.5	∞	0	4.17	
0.7	0.4	1.0	3.00	4.8	4.00	
0.7	0.4	1.5	2.31	19.8	5.95	This is a good design.
0.7	0.4	2.0	2.55	18.7	5.85	
0.7	0.4	2.5	2.60	25.7	6.30	
0.7	0.2	0.5	∞	0	4.30	
0.7	0.2	1.0	2.07	31.4	6.00	
0.7	0.2	1.5	2.30	22.6	5.85	This is a good design.
0.7	0.2	2.0	2.60	21.8	6.00	
0.7	0.2	2.5	2.60	19.2	6.20	
0.4	0	0.5	∞	0	11.50	
0.4	0	1.0	∞	0	11.22	
0.4	0	1.5	∞	0	11.40	This is an example of an overdamped system with long settling time.
0.4	0	2.0	∞	0	7.25	
0.4	0	2.5	∞	0	3.25	

st to provide a completely linear ensator designed for critical or over-ed operation, then remove the non- computer and insert the linear ensator when the values of $\dot{e}$ and e uitable, providing dead-beat response s fashion.

ferring to the data presented in Fig. 9, the authors give some indication of variations in per-cent overshoot en-tered, and of the variations in period e transient oscillations?

ERENCES

ULTIPLE MODE OPERATION OF SERVOMECH-, D. McDonald. *Review of Scientific Instru-*, New York, N. Y., 1952.

PTIMIZATION OF NONLINEAR CONTROL SYS-BY MEANS OF NONLINEAR FEEDBACKS, R. S. vander, R. H. MacNeal. *AIEE Transactions*, I (*Applications and Industry*), vol. 72, Sept. pp. 262–72.

N EXPERIMENTAL TREATMENT OF NONLINEAR OMECHANISMS, R. S. Neiswander. *Ibid.*, vol. ov. 1956, pp. 308–16.

SERVO WITH LINEAR OPERATION IN A REGION T THE OPTIMUM DISCONTINUOUS SWITCHING CURVE, L. L. Rauch, R. M. Howe. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, New York, N. Y., Apr. 1956.

Michael Athanassiades: We wish to thank Dr. Thaler for his remarks and especially for bringing to our attention the fact that the magnitude of the steady-state error for the system of Fig. 4 will not equal $(1+a)$.

From a mathematical point of view the magnitude of the steady-state error for the phase-plane geometry of Fig. 4 depends on the initial magnitude of the step and/or ramp input to the system. Dr. Thaler points out that in a physical system the steady-state error will become zero since there exists a finite error in the system. This implies that the system trajectory will move from the "end point" to the origin of the phase plane. However, the error velocity is zero and thus the system trajectory will be on the error axis. But, since the system has zero velocity, an infinite time is required in order to reach the equilibrium state of zero error, unless infinite gain is available somewhere in the loop. However, if an infinite gain amplifier may be used it may be placed in front of the system saturating amplifier thus reducing the entire system to the familiar second-order "bang bang" case. This need for infinite gain is best demonstrated by considering the nonlinear computer of Fig. 13. The phase-plane geometry of Fig. 4 may be obtained from the computer of Fig. 13 if the characteristics of the limiter shown, with maximum output $\pm(1+a)$, are replaced by a perfect relay-type nonlinearity, with the same maximum output level, which would deliver the signal $\phi(\dot{e}) = (\text{sgn } \dot{e})\, \dot{e}^2(1+a)$.

Table II is included in order to give some indication of the transient response of the system with the nonlinear computer shown in Fig. 8. In the table the transient rise time is defined as the time required for the error to reach its zero value for the first time. Thus, an infinite rise time implies that the step response of the system was such that the error approached zero asymptotically without any overshoot.

Stability Analysis of Dual-Mode Servomechanisms

JOHN E. GIBSON
MEMBER AIEE

EUGENE S. McVEY
NONMEMBER AIEE

opsis: Dual-mode servomechanisms e several advantages such as light ght and simplicity, but a serious study the stability problems encountered in h devices has not appeared in the rature. An analytical method is pre-ted here for determining the stability dual-mode servomechanisms by an ension of the describing function method analysis. The method is illustrated h examples, and laboratory data are sented. Analysis by describing functions ws that a dual-mode system may be stable even though its linear and relay des of operation are stable as systems themselves, and the mode-switching cuitry is perfect. It is also shown that certain mode boundary shapes the cribing function is a function of fre-ency as well as amplitude.

A DUAL-MODE servomechanism is an automatic control system that nbines two separate and distinct modes operation in a manner that makes the r-all system superior to a system lizing either of the modes alone. The rticular mode of operation at any time elected in a manner to make use of the t features of each mode under the given erating conditions and, if possible, to minate the undesirable features of each de. The dual-mode system is the simplest and perhaps the most practical of the general class of systems called "multiple mode systems." The advantages of dual-mode systems are well known, and recorded in the literature. Perhaps the greatest deficiency in dual-mode theory is the lack of a method for determining system stability.

The analytical state of the art for dual-mode systems is to analyze each mode independently. This approach appears quite reasonable and yields much performance information. In addition, for the case of the linear and relay mode combination, because these types of operation are better understood than any others, a large body of information exists concerning each mode of operation. However, this analytical method suffers a critical deficiency in that it yields no information on the stability of the over-all system. The fact that the linear and relay modes are stable as individual systems gives no assurance that the dual-mode system is stable even if the mode-determining and switching circuitry is perfect. It is the purpose of this article to present a method for determining the stability of a dual-mode system.

McDonald[1] has described a dual-mode servomechanism which acts as a relay system for large values of error and as a linear system for small values of error. This is the dual-mode combination of most interest because it requires a power amplifier which is only a fraction of the size and weight of the power amplifier required in a linear system with the same actuator. In addition, this combination yields advantages such as simplicity, economy, and reliability, and has the advantages of a linear system for small values of error and the advantages of a relay system for large values of error.

Fig. 1(A) shows a possible configuration for the general system discussed in this article. The configuration acts as a linear system when the mode switch is in the de-energized position as shown, and as a relay system when the mode switch is energized. The mode switch is usually a relay but the word "switch" is used to distinguish it from the polarized relay referred to here as the power relay, which is used as an amplifier in the relay portion of the system. The simplest way to instrument the switching-boundary circuit for the relay mode amplifier is to utilize the input characteristics of the power relay. For simplicity, this method will be used although the stability analysis is applicable to more complicated switching boundaries. Likewise, the simplest way

Paper 60-117, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 6, 1959; made available for printing December 16, 1959.

JOHN E. GIBSON and EUGENE S. McVEY are with Purdue University, Lafayette, Ind.

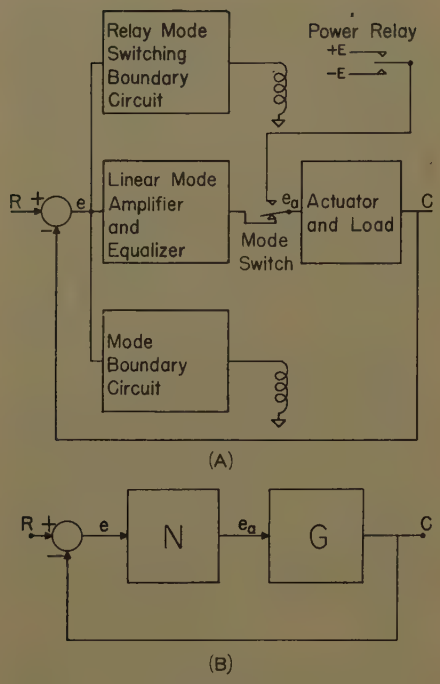


Fig. 1. Dual-mode system

A—Configuration
B—Block diagram

to instrument the mode boundary circuit is to utilize the input characteristics of the mode switch. When the mode switch is a relay, its hysteresis and dead band establish the mode boundary. In this article the describing function analysis is used to analyze the stability of dual-mode systems with both perfect relays and practical relays having hysteresis and dead band.

The problem of mode boundary instability has apparently not been the subject of previous work in the literature. It is shown here that this can become the critical factor in the successful operation of the over-all system.

Nomenclature

α = displacement of center of power relay voltage from a time reference point
β = one-half of power relay voltage "on time" (e.g., see Fig. 4)
Δ = relay dead band
C = system output
D_1 = fundamental component of voltage applied to motor armature
e = system error signal
e_a = voltage applied to motor armature
e_m = mode boundary magnitude; magnitude of system error or modified error at which system switches from one mode to other
E = magnitude of voltage from power relay
E_1 = peak amplitude of system error signal
G = transfer function for linear part of system
h = relay hysteresis
K = gain of linear mode amplifier
K_{eq} = equivalent gain of nonlinearity; describing function
N = symbol for nonlinear transfer function
R = system input

Stability Analysis

The independent work of Goldfarb,[2] Kochenburger,[3] and Tustin[4] has established the describing function method of analysis for the approximate determination of nonlinear system stability. The describing function or equivalent gain, as it is also called, is used as a quasi-linear transfer function for the nonlinear element. If an assumption is made that the linear portion of the system acts as a low-pass filter to the extent that harmonics produced by the nonlinearity can be neglected, then for a sinusoidal input the equivalent gain is the ratio of the fundamental component of the nonlinearity output to the input. The fundamental component is obtained by expressing the output of the nonlinearity as a Fourier series. Fig. 1(B) is a typical block diagram for a nonlinear system where N represents a nonlinearity and G represents the linear part of the system. Generally, G will contain a motor or other power element which acts as a low-pass filter in the frequency range of interest. The essential difference between the analysis of a relay system and a dual-mode system is that in the case of the dual-mode system the linear mode amplifier, the relay mode amplifier (the power relay and its associated switching boundary circuitry), and the mode bundary circuitry comprise the nonlinearity, whereas in the relay system one is concerned only with a relay amplifier.

The characteristic of the nonlinearity for a dual-mode system is obtained by considering the physical action of the circuitry involved. For example, consider a system with a perfect mode switch which switches from one mode to the other when the magnitude of the system error is equal to e_m, as shown in Fig. 2. The plot of e_a versus e in Fig. 3 represents the action of the circuitry from the error point to the motor armature and is the nonlinearity for this system. For values of error with absolute magnitude less than e_m the system operates in the linear mode and the slope of the nonlinearity is equal to K, the gain of the linear mode amplifier, and for values of error with absolute magnitude greater than e_m the system operates in the relay mode with an output of the power relay equal to E in magnitude and with the sign of the error signal.

The output e_a of the nonlinear part of the system may be determined for a sinusoidal input by analysis of the nonlinear characteristic. Of course, if desired, one may determine this directly

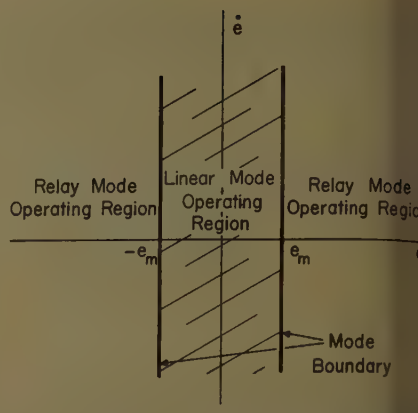


Fig. 2. Phase plane of error showing simp mode boundary

from physical considerations without d termining the nonlinear characteristic the system. The output corresponding the nonlinearity of Fig. 3 is shown in Fi 4. The fundamental component of th voltage is determined and used in th expression for the equivalent gain.

After obtaining the equivalent gain for dual-mode system the procedure for dete mining the stability of the system identical to the procedure used for simp relay systems. Referring to Fig. 1(B the quasi-equation

$$\frac{C}{R}=\frac{NG}{1+NG} \qquad (1$$

can be written. This expression ha meaning if N is adequately described b an equivalent gain function. The syste is unstable if the denominator is equal t zero, i.e.,

$$1+K_{eq}G=0 \qquad (2$$

or

$$G=-\frac{1}{K_{eq}} \qquad (3$$

The equivalent gain describes N whe the conditions under which it was derive

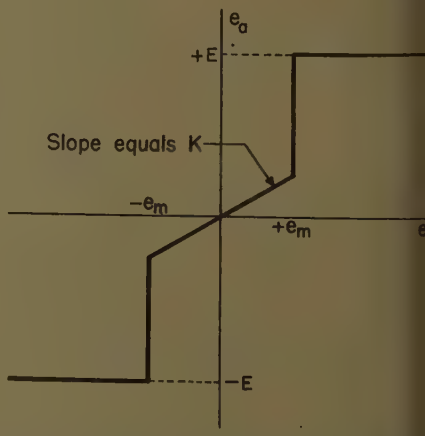
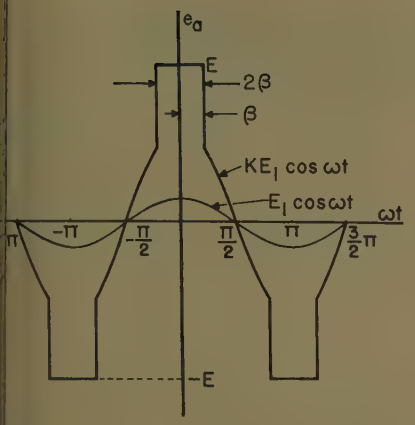


Fig. 3. Voltage applied to motor armatu versus error voltage: nonlinear representatio of dual-mode amplifier characteristics



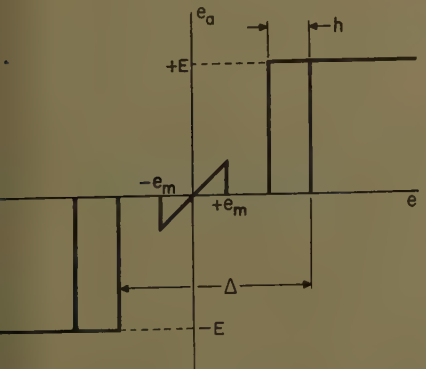
4. Motor armature voltage for system h perfect mode switch and mode boundary shown in Fig. 3

present. These conditions are adeately satisfied if the system is in free illation and if G contains sufficient lowss filtering that the error signal can be sumed to be sinusoidal. The stability a system can be determined by plotting and $-1/K_{eq}$ on a polar plot. If the o curves intersect, the identity of equan 3 is satisfied and the system is capable oscillating. Like the case of a system th dead band, a disturbance is necessary fore the system will start oscillating. he amplitude and frequency of oscillaon can be determined directly from the lar plot.

The stability analysis method for dualode systems will now be illustrated in tail with examples.

XAMPLE 1: PERFECT MODE SWITCH

Dual-mode system stability will be rst illustrated using a perfect mode vitch. The system configuration of Fig. will be used and the mode-boundary rcuit will be assumed to have a transfer unity. The mode boundary is shown in ig. 2. The value of e_m will be chosen to e greater than the operating voltage of he power relay. This makes e_a inde-



ig. 5. Nonlinear characteristic for system ontaining perfect mode switch in which e_m less than the power relay operating voltage

pendent of the characteristics of the power relay since, under this condition, the power relay switches to the correct position before the mode switch puts the system into the relay mode of operation. The mode switch also switches the system back to the linear mode before the power relay changes to a different position. If e_m is not chosen greater than the power relay operating voltage, an inactive zone will be present. This inactive region is illustrated in Fig. 5 for a power relay with hysteresis and dead band, and has a width of $[(\Delta+h)/2-e_m]$ when the system switches from the linear to the relay mode and a width of $[(\Delta-h)/2-e_m]$ when the system switches from the relay to the linear mode. Since one purpose of the dual-mode system is to eliminate the inactive region of operation present in certain relay systems for small values of error, this minimum value of e_m restriction is justified and should be observed in design.

The voltage applied to the motor armature (actuator) for a sinusoidal error signal may be determined by application of the nonlinear representation of the dual-mode amplifier characteristics, Fig. 3, and yields the waveform of the voltage e_a as shown in Fig. 4. The assumed error signal is

$$e = E_1 \cos \omega t \tag{4}$$

The fundamental term of the Fourier series for e_a is now determined for use in the equivalent gain representation of the transfer from error to the motor armature. The mode relay switches when

$$|e| = e_m \tag{5}$$

Replacing e by its value from equation 4 yields

$$|E_1 \cos \omega t| = e_m \tag{6}$$

From Fig. 4, β is one half of the power relay ON period and is the vaue of ωt at which switching occurs. From equation 6, β becomes

$$\beta = \cos^{-1} e_m/E_1 \tag{7}$$

Due to the symmetry of e_a, only the a_1 coefficient of the Fourier series needs to be calculated. It is

$$a_1 = \frac{1}{\pi}\,[4E \sin \beta + KE_1(\pi - 2\beta - \sin 2\beta)] \tag{8}$$

The equivalent gain is then

$$K_{eq} = \frac{1}{\pi E_1}\,[4E \sin \beta + KE_1(\pi - 2\beta - \sin 2\beta)] \tag{9}$$

A convenient check on the correctness of the equivalent gain is to note that it approaches K, the gain of the linear mode amplifier, as β approaches zero, and it also approaches the describing function for a relay system with a perfect relay as β approaches $\pi/2$.

System stability may now be determined by application of equation 3. Calculated data are presented in Fig. 6 for a system having the following assumed values:

$e_m = 10$ volts
$K = 2$
$E = 100$ volts

Two different linear functions, G_1 and G_2, are assumed to illustrate stable and unstable dual-mode systems respectively. The assumed second-order transfer function G_1 will result in a stable system independent of the gain or time constant of the system since its phase shift is never great enough to cause an intersection.

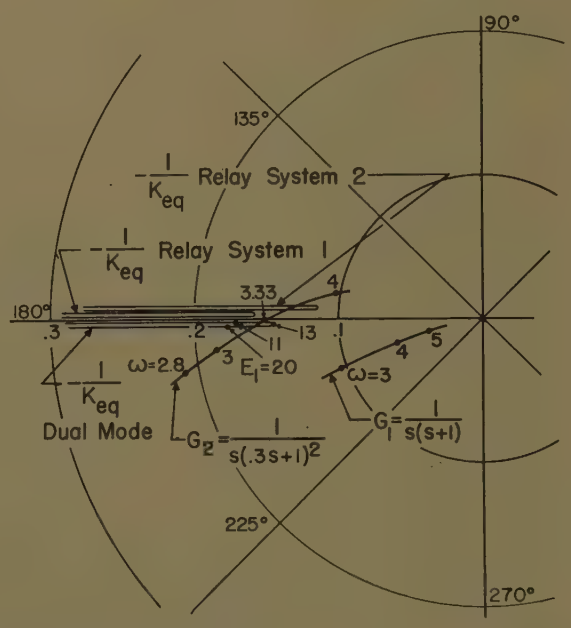


Fig. 6. Polar plot for dual-mode system with perfect mode switch, two relay systems also shown

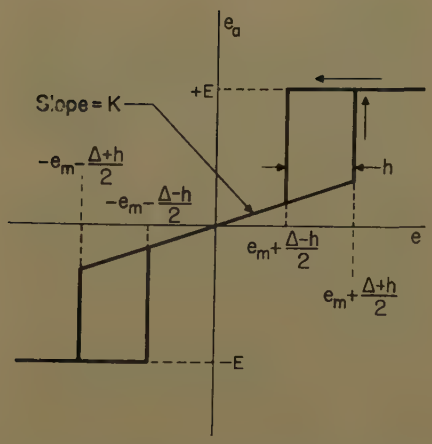


Fig. 7. Nonlinear characteristic for system with hysteresis and dead band at mode boundary

The linear transfer G_2 intersects $-1/K_{eq}$ twice, resulting in an unstable system with a predicted frequency of oscillation of 3.33 rad/sec (radians per second) at a peak amplitude of 15 volts for the convergent equilibrium intersection.

The open-loop transfer function of the linear mode as a system by itself using G_2 is as follows:

$$A=\frac{2}{S(0.3S+1)^2} \tag{10}$$

Application of the Hurwitz stability criterion readily establishes that the linear mode is stable as a system by itself. A simple relay system using a relay with no hysteresis and a value of dead band equal to 20 volts would result in a stable system by itself when the linear part of its system is G_2. This system is labeled Relay System 1 in Fig. 6; it could be combined with G_2 to make the dual-mode system analyzed above. Thus, it is possible to have a dual-mode system which is composed of a relay mode and a linear mode which are stable as systems by themselves, but which are unstable when combined into a dual-mode system. Relay system 2 is equivalent to relay system 1 except the power relay dead band is 16 volts. This system is unstable with the linear transfer G_2 as a relay system by itself, but as a dual-mode system with the values assumed above, except for $K = 1$, it is stable. Thus, the well-known fact that a relay system can be stabilized by the addition of a linear mode is demonstrated by this stability analysis.

The values chosen in this example are arbitrary. However, it can easily be proved that a dual-mode system with a perfect mode switch is in general more prone to oscillate than a relay system with a power relay dead band equal to twice the amplitude of e_m (the operating voltage of the power relay is equal to e_m, which is the case when the linear mode is made as small as possible without introducing an inactive region of operation). This can be demonstrated by showing that the equivalent gain for the dual-mode system is always greater than the equivalent gain of the relay system:

$$\frac{1}{\pi E_1}\left[4E\sin\beta+E_1K(\pi-2\beta-\sin 2\beta)\right]>\frac{4E\sin\beta}{\pi E_1} \tag{11}$$

$$\pi-2\beta-\sin 2\beta>0 \tag{12}$$

For a value of β equal to $\pi/2$ the "greater than" condition does not hold. However, this is a limiting condition for which the dual-mode system degenerates into a single mode relay system; hence, the validity of equation 12 is not effected. The inequality of equation 12 can be conveniently proved with the mean value theorem from the calculus.

A physical explanation of the fact that a dual-mode system may tend toward instability, more than a relay system which uses the same linear transfer G, is that the energy from the linear mode increases the equivalent gain of the dual-mode system over that of the relay system.

Example 2: Hysteresis and Dead Band at the Mode Boundary

In this example the characteristics of a practical mode switch are assumed. Hysteresis and dead band are taken into account. The input impedance to the relay is assumed to be resistive over the frequency range of interest. Experimental data justify this assumption.

The system of Fig. 1 is assumed. The system from e to e_a can be represented by the nonlinearity shown in Fig. 7. Application of Fig. 7 with a sinusoidal error signal yields the waveform e_a shown in Fig. 8(A).

The mode switch changes the system from the linear to the relay mode when

$$|e|=e_m+\frac{\Delta+h}{2} \tag{13}$$

and it follows that

$$\beta-\alpha=\cos^{-1}\frac{2e_m+\Delta+h}{2E_1} \tag{14}$$

The mode switch switches from the relay mode to the linear mode when

$$|e|=e_m+\frac{\Delta-h}{2} \tag{15}$$

and it follows that

$$\beta+\alpha=\cos^{-1}\frac{2e_m+\Delta-h}{2E_1} \tag{16}$$

Solving equations 14 and 16 simultaneously yields

$$\alpha=\frac{1}{2}\left[\cos^{-1}\frac{2e_m+\Delta-h}{2E_1}-\cos^{-1}\frac{2e_m+\Delta+h}{2E_1}\right] \tag{17}$$

$$\beta=\frac{1}{2}\left[\cos^{-1}\frac{2e_m+\Delta-h}{2E_1}+\cos^{-1}\frac{2e_m+\Delta+h}{2E_1}\right] \tag{18}$$

To determine in this case the fundamental component of e_a, it is necessary to calculate both the a_1 and b_1 terms of the Fourier series which represents e_a.

The fundamental component of e_a is

$$D_1=a_1\cos\omega t+b_1\sin\omega t \tag{19}$$

$$=r\cos(\omega t-\theta) \tag{20}$$

where

$$r=(a_1^2+b_1^2)^{1/2} \tag{21}$$

$$\theta=\tan^{-1}b_1/a_1 \tag{22}$$

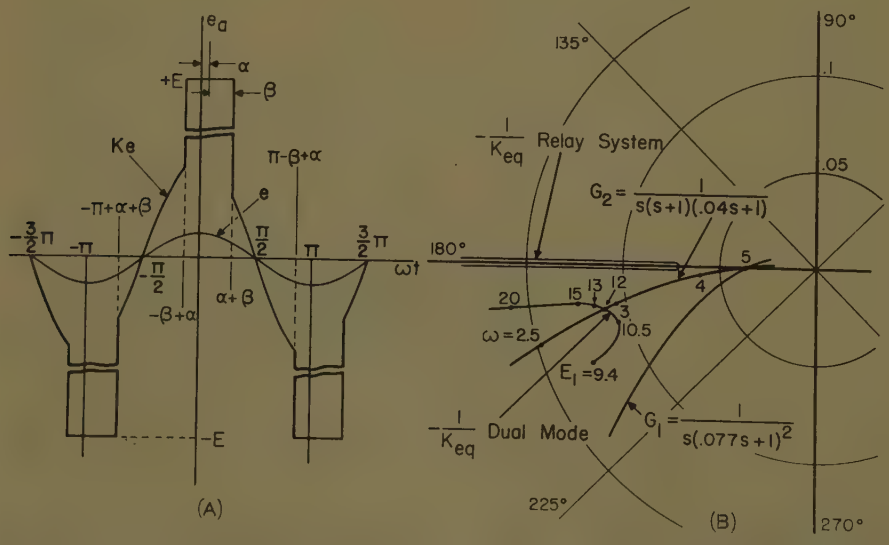


Fig. 8. A—Motor armature voltage for system with hysteresis and dead band at mode Boundary. B—Polar plot

Table I. Data for Example 2

	Mode Switch	Power Relay
Hysteresis.....	5.0 volts	.....negligible
Dead band.....	13.8 volts	.....4.4 volts
Input resistance.....	152 ohms	
Input inductance....	0.38 henry*	

* Approproximate.

The equivalent gain becomes

$$K_{eq}=\frac{r}{E_1}\angle -\theta \tag{23}$$

where

$$r=\left(\frac{1}{\pi^2}[KE_1(\pi-2\beta)+4E\cos\alpha\sin\beta]^2+\frac{1}{\pi^2}\left\{4E\sin\alpha\sin\beta+KE_1[\sin^2(\alpha-\beta)-\sin^2(\alpha+\beta)]\right\}^2\right)^{1/2} \tag{24}$$

$$\theta=\tan^{-1}\frac{4E\sin\alpha\sin\beta+KE_1[\sin^2(\alpha-\beta)-\sin^2(\alpha+\beta)]}{KE_1(\pi-2\beta)+4E\cos\alpha\sin\beta} \tag{25}$$

It is of interest to note that this equivalent gain reduces to the equivalent gain for a relay servo if K and e_m are set equal to zero and h and Δ are taken as the characteristics of the power relay instead of those of the mode switch.

A system using production relays has been tested in the laboratory. An analog computer was used to simulate the linear portion of the system. The relays have the characteristics shown in Table I. The value of e_m was set equal to zero; thus the mode switch established the mode boundary. The power relay characteristics are unimportant since its operating voltage is less than that of the mode switch. The system has $K=2$ and $E=100$ volts.

Analytical data for the system are shown in Fig. 8(B). The sysem is stable with the linear transfer G_1 and unstable with G_2. The predicted oscillation has an amplitude of 12 volts peak and a frequency of 2.95 rad/sec. Fig. 9 shows experimental data for the system. The observed oscillation had an amplitude of 13 volts at a frequency of 2.73 rad/sec, which is considered to be within experimental accuracy, and to be the accuracy one may expect from the describing funtion analysis without using correction terms.

The system is stable until the error is made large enough for the system to switch into the relay mode. Once this is done, it oscillates as shown in Fig. 9. Sinusoidal and step inputs of different amplitude were used to start the oscillations. It was found that the system oscillated about the value $C=R$ for step inputs; hence, no d-c component was present in the error signal which is in agreement with the analytical assumptions.

An analysis of the relay mode as a system by itself is included in Fig. 8(B). The analysis predicts that it would be stable with the linear transfer G_2.

Example 3: Perfect Mode Switch and Power Relay, e+ė Mode Boundary

This example illustrates the use of the stability analysis method with a mode boundary different from that used in the previous examples. The system switches from one mode to the other at the mode boundary defined by

$$|e+\dot{e}|=e_m \tag{26}$$

If e is assumed to be

$$e=E_1\sin\omega t \tag{27}$$

then

$$e+\dot{e}=\sqrt{E_1^2+E_1^2\omega^2}\cos(\omega_t-\psi) \tag{28}$$

where

$$\psi=\tan^{-1}\frac{1}{\omega} \tag{29}$$

The voltage e_a is shown in Fig. 10(A) with the voltage $e+\dot{e}$ used as a reference. The equivalent gain is determined in the same manner as in the previous examples and is

$$K_{eq}=\frac{r}{E_1}\angle\phi-\theta \tag{30}$$

where

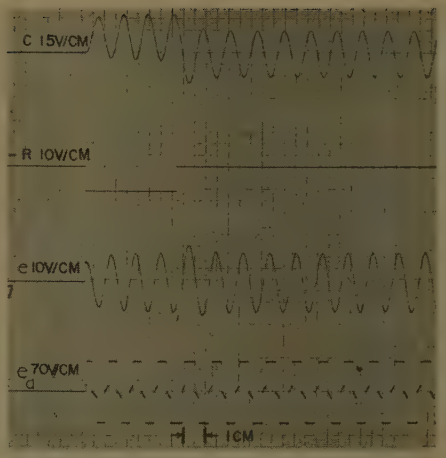


Fig. 9. Experimental waveforms for system of example 2. One centimeter=2 sec

$$r=\left\{\left[\frac{4E}{\pi}\sin\beta+\frac{KE_1}{\pi}(\pi-2\beta+\sin 2\beta)\cos\phi\right]^2+\left[\frac{KE_1}{\pi}(\pi-2\beta+\sin 2\beta)\sin\phi\right]^2\right\}^{1/2} \tag{31}$$

$$\beta=\cos^{-1}\frac{e_m}{E_1(1+\omega^2)^{1/2}} \tag{32}$$

$$\theta=\tan^{-1}\frac{KE_1(\pi-2\beta+\sin 2\beta)\sin\phi}{4E_1\sin\beta+KE_1(\pi-2\beta+\sin 2\beta)\cos\phi} \tag{33}$$

$$\phi=90°-\psi \tag{34}$$

In this example the oscillation frequency ω is a parameter of the equivalent gain. This necessitates plotting a family of equivalent gain curves, one for each value of ω. Instability requires not only that the negative reciprocal of the equivalent gain curve intersect the linear transfer part of the system but, in addi-

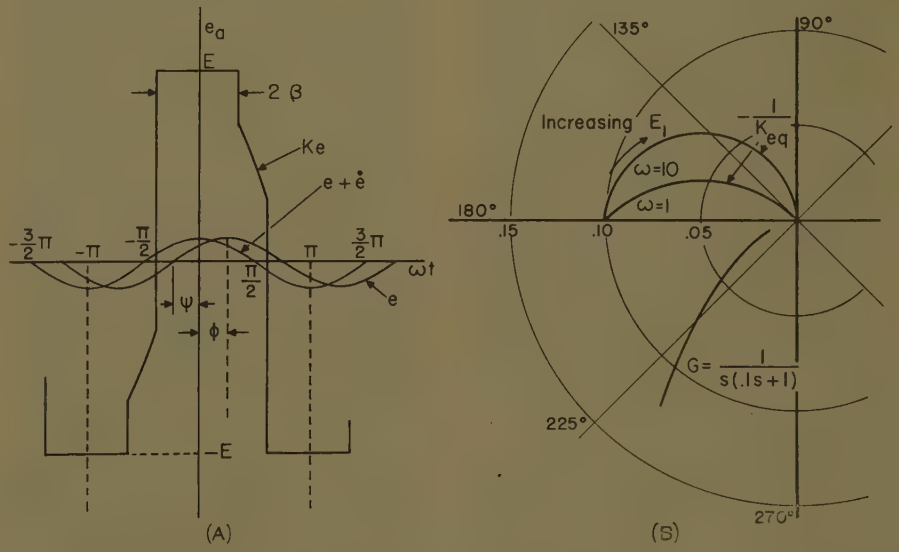


Fig. 10. A—Motor armature voltage for system with perfect mode switch, perfect power relay, and e+ė mode boundary. B—Polar plot

tion, that it must also intersect the linear curve at the frequency ω.

Curves are plotted in Fig. 10(B) for a system where

$E = 100$ volts
$e_m = 1$ volt
$K = 10$

It may be seen from the figure that the system is stable.

Conclusions

The stability of dual-mode servomechanisms has been investigated by means of the describing function method, and the method illustrated with examples and supported with experimental data.

It has been established analytically that an unstable dual-mode system is possible even though the linear and relay modes are stable as systems by themselves and a perfect mode switch assumed.

It has been established that the mode switch (or mode-determining circuitry) characteristics play a predominant role in system stability; in fact, the characteristics of the power relay do not enter into stability considerations other than to establish the minimum size of the mode boundary in the examples presented. There are other system configurations where this rule does not apply. For example, if the voltage to the power relay is routed through the mode switch contacts the characteristics of both relays must be taken into account, or if an imperfect relay is assumed in example 3 the power relay characteristic would enter directly into the analysis. However, the important point illustrated here is the necessity of taking into account the mode-determining elements system stability, and how this is don Finally, it has been shown that for c tain mode boundary shapes the resultin describing functions are frequency sens tive.

References

1. Multiple Mode Operation of Serv mechanisms, **D. McDonald.** *Review of Scienti Instruments*, New York, N. Y., vol. 23, 1952, p 22–30.

2. On Some Nonlinear Phenomena in Reg latory Systems, **L. C. Goldfarb.** *Automatika Telemekhanika*, Moscow, U.S.S.R., vol. 8, no. Sept.-Oct. 1947, pp. 375–83.

3. A Frequency Response Method for An lyzing and Synthesizing Contactor Serv mechanisms, **R. J. Kochenburger.** *AIEE Tra actions*, vol. 69, pt. I, 1950, pp. 270–83.

4. The Effects of Backlash and of Spee development Friction on the Stability Closed-Cycle Control Systems, **A. Tusti** *Journal*, Institution of Electrical Engineer London, England, vol. 94, pt. II-A, no. 1, 1947.

Discussion

Michael Athanassiades (University of California, Berkeley, Calif.): I should like to point out that the procedure presented in this paper for determining the stability conditions for a nonlinear system was described in a book by Smith in 1958.[1] The identical procedure and many examples of application appear in chapters 11 through 14, although the graphical methods employed are different. Thus any nonlinear control system containing a nonlinearity N and a linear member G, is analyzed by obtaining the equivalent gain and phase function of the nonlinear element using the describing function method and determining the intersection of two curves, according to equation 3.

It must be emphasized that this general procedure is completely general and may be used for any system containing nonlinear elements as long as the system, naturally or by means of block diagram substitutions, may be separated into a nonlinear element N followed by a linear element G, connected in feedback. In addition, the N is not limited to relay-type input-output characteristics, but also may result from backlash, velocity hysteresis, coulomb friction and stiction, or any other combination of these nonlinearities.

Identical results may be obtained through the use of the L-plane, instead of the polar plot used by the authors. The L-plane, described in the reference, is a logarithmic conformal mapping of the s-plane. The co-ordinates of the L-plane are degrees of phase-lag angle as the abscissa and decibels of gain as the ordinate. This mapping results in a scale relation, in such a way that 6 decibels must equal 39.6 degrees of phase.

When the equivalent gain and phase of nonlinearity and the frequency transference of the linear system are plotted on the same L-plane, the intersection or intersections of these two curves give the frequency of oscillation in the absence of input, the oscillation amplitude, and the phase of the closed-loop nonlinear system.

The advantage of the L-plane method of analysis is that, using simple distances measured on the L-plane, one may predict quite accurately the transient response of a nonlinear system. For example, if a sine wave is used as the input signal whose frequency is different from the one indicated by the curve intersection, and then removed, the system has a transient which may be calculated from the L-plane data. Moreover, it is very easy to design compensating networks and/or adjust the system gain using the L-plane plot, so that the general system performance may improve and meet certain given specifications.

Reference

1. Feedback Control Systems (book), **Otto J. M. Smith.** McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

Franz B. Tuteur (Yale University, New Haven, Conn.): The question of the stability of dual-mode servomechanisms is one that has needed investigation for some time, and the authors are to be congratulated for presenting this first analysis of an important problem. Since the apparent purpose of the authors was primarily to point out the general problem of the possible instability of even simple systems of this type, one cannot find fault with them for not considering every possibility. However, it is suggested that in future investigations they address themselves to the matter of drift in the linear amplifier, and the effect of the charge that may be built up on the capacitors in the equalizing networks of the linear amplifier by large errors. It has been found in our own investigations on this problem that these factors may also result in instability or, at best, very slowly decaying oscillations.

A further point that should perhaps be made is that the power relay in many relay-controlled systems is controlled by a computer which computes a generally nonlinear switching boundary that depends on the error and its time derivatives. As a matter of fact, the need for a dual-mode system arises in the first place from the fact that relay systems designed for opt mum transient response are unstable ne zero error. It seems to us that, dependin on the exact form of the switching boundar for the power relay, it is quite possible fo waveshape of the motor voltage (Fig. 4) t look quite different, with several reversal of the power relay occurring between th instants at which the mode switchin occurs. The describing functions of th over-all relay system would therefore b quite different from that computed by th authors.

J. E. Gibson and **E. S. McVey:** The author wish to thank the discussers for takin time to prepare written discussions. It i felt that such discussions enhance the valu of a paper for the reader.

It is our impression that perhaps Mr Athanassiades has mistaken the purpos of the paper. The authors are awar that the describing function method o analysis is presented by Smith, as well a by others. The method is well know and it is felt that it was necessary only t refer to Goldfarb, Kochenburger, an Tustin as the authors of the first papers o the subject. It was not the purpose of th paper to illustrate the describing functio method of analysis, but rather to use thi tool to analyze dual-mode systems. T our knowledge, dual-mode systems hav not previously received this attention.

While it might be said that Goldfarb' paper contains the germ of all of th subsequent work concerned with describin function analysis, this hardly preclude the usefulness of all later books and papers Perhaps the discusser merely feels that th stability problem in dual-mode systems i trivial. With this position it is difficul to argue, of course, since it depends o each individual's insight and intuition The authors naturally feel otherwise Indeed, they feel supported in their con tention that all is not obvious and intuitive on the basis of the discusser's reference In four separate places[1] Smith implie that it is sufficient to examine the stabilit of the linear mode alone for small signal

that the analysis of each mode separately r stability is sufficient. The present per proves that these intuitive conclu- ons are not true.

Concerning the use of the L-plane in is analysis, it is true that this alternate proach could be used. It is felt that no rther comments are necessary on the plane method since this is not the subject the paper. However, this is not to be nstrued to mean that the authors agree ith all of the rather sweeping claims made ncerning the evaluation of dynamic sponse and system synthesis by this chnique.

The authors wish to thank Dr. Tuteur r his constructive suggestions. We have t investigated the transient phenomena entioned but agree that work needs to be done in this area. It is true that the stability analysis becomes more complicated when switching boundaries are introduced to optimize the nonlinear mode of operation; however, the method of analysis presented here is applicable. Of course, as Dr. Tuteur points out, one of the purposes of employing a dual-mode system is to improve the small-signal operation of the system, and a second purpose is to allow the simplification of the switching boundaries required for approximately optimum performance. Work is presently under way on a study of these simplified switching boundaries and on optimum mode boundaries in dual-mode systems. Other work being done in this area by the authors is the study of jump and subharmonic phenomena which have been observed when a dual-mode system is driven by a sinusoidal input.

The authors are pleased to note the current increase in interest in dual-mode systems.[2] We feel that because of their several practical advantages over pure linear systems on the one hand and relay systems on the other, dual-mode systems will come into wider application in power servomechanisms. It is somewhat more difficult to justify their complexity in instrument servo applications.

References

1. See reference 1 of the **Athanassiades** discussion pp. 508, 536, 592, and 602.

2. Dual-Mode Relay Servos, **R. N. Buland, N. Furumoto.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, 1959 (Jan. 1960 section), pp. 405–11.

Forward Voltage Drop and Power Loss in Silicon Rectifiers

WERNER LUFT
ASSOCIATE MEMBER AIEE

Synopsis: The instantaneous forward voltage drop of silicon diodes of different sizes and processes is investigated over four magnitudes of current density. The temperature influence is determined. An equation relating the instantaneous forward voltage drop to current density and junction temperature is given. An expression of the average forward voltage drop and power loss as function of current density and conduction angle is developed.

TO EVALUATE the current-carrying capacity of a semiconductor diode, it is necessary to know the power losses within the diode as a function of the current. The power losses are forward conduction losses and reverse blocking losses. The former are generally of much greater magnitude than the latter, which depend on the reverse voltage and leakage current. The leakage currents are usually held to such limits that the blocking losses are the same for diodes of different voltage grades. Therefore, only the forward conduction losses will be discussed.

The instantaneous forward power loss is a function of instantaneous current, junction area, kind of junction, and operating temperature. The average power loss is, therefore, also a function of the current waveshape. For an a-c power supply and a load giving sinusoidal current through the diode, the average power loss will thus vary with the conduction angle.

The average power loss can be determined from the instantaneous current versus forward voltage-drop characteristic of the rectifier. This can be done graphically for any current waveshape, but a mathematical method is preferable. For the most common case, sinusoidal voltage supply, such a method is presented here.

The temperature influence on the forward voltage drop has been investigated and is considered in the developed equations.

Nomenclature

$I_{\phi_1\phi_2}$ = average current over 2π for conduction from ϕ_1 to ϕ_2, amperes (amp)
s = instantaneous current density, amperes/centimeter^2 (amp/cm^2)
$\hat{S}$ = peak current density, amp/cm^2
$S_{\phi_1\phi_2}$ = average current density over 2π for conduction from ϕ_1 to ϕ_2, amp/cm^2
e = instantaneous forward voltage drop, volts
$E_{\phi_1\phi_2}$ = average forward voltage drop over 2π for conduction from ϕ_1 to ϕ_2, volts
ϕ_1 = angle at start of conduction, radians
ϕ_2 = angle at end of conduction, radians
α = variable angle, radians
f_1 = constant, volts cm^{2g} C/amp^g
$f = f_1/T$ resistance (exponential), volts cm^{2g}/amp^g
g_1, g_2 = exponents
$g = (g_1 T - g_2)$, exponent
h_1 = resistance (exponential), volts cm^{2m}/amp^m
β = temperature coefficient, K^{-1}
$h = h_1 (1+\beta T)$, resistance (exponential), volts cm^{2m}/amp^m
m = exponent
F = area of junction, cm^2
$P_{\phi_2\phi_2}$ = average power loss over 2π for conduction from ϕ_1 to ϕ_2, watts
T = junction temperature, degree Kelvin (K)
$\psi_{\phi_1\phi_2}$ = form factor

Forward Voltage Drop

The forward voltage drop as function of current was measured with an oscilloscope in a circuit as shown in Fig. 1. To limit the temperature rise of the junction for current densities up to 3,000 amp/cm^2, a capacitor discharge of approximately 0.5 millisecond duration was used. The temperature rise caused by the current pulse was calculated by a method described by Diebold.[1] Tests were performed on alloyed-junction diodes of several sizes and on diffused-junction diodes of one size.

Variations in the voltage drop for diodes of the same type and size, and at the same temperature, depend mainly on three factors: 1. variation in thickness of the silicon wafer, 2. variation in the junction area, and 3. variations in the ohmic con-

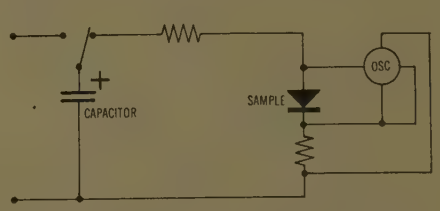


Fig. 1. Circuit for forward voltage drop measurements

Paper **60-34**, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 31, 1959; made available for printing December 4, 1959.

Werner Luft is with the International Rectifier Corporation, El Segundo, Calif.

The author wishes to acknowledge the helpful suggestions and criticisms in regard to this paper given by Mr. E. J. Diebold.

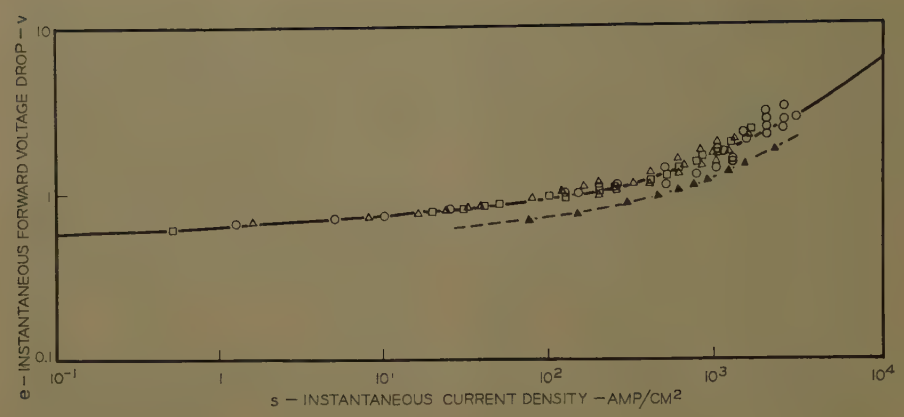


Fig. 2. Instantaneous forward voltage dro versus instantaneous current density f alloyed-junction diodes and for diffuse junction diodes

Symbol	F(cm
Alloyed-Junction Diodes	
○	0 39
△	1.26
□	1.98
Diffused-Junction Diodes	
▲	0 06

T = 323 K

tacts within the rectifier. During the manufacture of the diodes the first factor is controlled within 0.002 inch. For diodes of the area-alloyed type, the junction area is determined by the size of the alloying material and the alloying time. The area of the alloying material is held within 3%. The main variations of the forward voltage drop are thus caused by variation in the ohmic contact. The contact may be unsatisfactory between the silicon and the materials to which it is bonded; between this material and the base, or where the terminal lead enters the diode and where it is attached to the junction.

The higher forward voltage drop caused by poor ohmic contact is most easily seen at high current densities. To eliminate such diodes, production tests are made on all diodes with short pulses of high current. The current-voltage characteristic is viewed on an oscilloscope, and diodes having a drop above a certain value are rejected.

Fig. 2 illustrates diode forward characteristics plotted as logarithm of instantaneous forward voltage drop versus logarithm of instantaneous current density in the junction. Both alloyed-junction and diffused-junction diodes are represented. The junction areas for the alloyed diodes, which are all of aluminum-silicon-type alloy, are 0.396, 1.263, and 1.98 cm² and for the diffused diodes 0.067 cm². The junction temperature during the measurements was held constant at approximately 50 degrees centigrade (C).

It is seen that the alloyed-junction diodes follow closely the same characteristic curve for all investigated sizes of junction area. Data from other investigators indicate that the shape of the curve is maintained to about 10^{-2} to 10^{-3} amp/cm² before an inflexion occurs.[2]

The diffused-junction diodes have the same general characteristic, but the voltage drop at each current density is lower as shown by the dashed line in Fig. 2. The spread in test points reflects the variation in forward characteristics attributable to production tolerances and depends, as mentioned, mainly on the differences of ohmic contacts within the diode.

Analytical Expression

Many trials were made to find a simple equation which would closely match the test data for a wide range of current densities. The result is the analytical expression

$$e = f(s)^q + h(s)^m \qquad (1)$$

This equation has the advantage that the forward voltage drop goes to zero when the current density goes to zero, and does not assume a positive value as in equations containing a constant term or a negative value as in equations including a term in which the logarithm of the current occurs.

Numerical values of equation 1, giving the best matching to the test points, are for the alloyed-junction diodes

$$e = 0.645(s)^{0.561} + 1.45(10^{-3})(s)^{0.9832}$$

and for the diffused-junction diodes

$$e = 0.48(s)^{0.0682} + 8(10^{-4})(s)^{0.918}$$

How close matching has been obtaine for current densities between 0.1 and 10 000 amp/cm² can be seen by the soli line in Fig. 2.

The relative magnitude of the two term varies greatly with current density. A 10 amp/cm² the second term is only abou 1% of the first; at approximately 1,50 amp/cm² the two terms assume the sam magnitude and at still higher curren densities the second term becomes dom inant.

The physical meaning of equation 1 that the diode is equivalent to two non linear resistors in series. The degre of nonlinearity is described by the expo nents, which theoretically cannot excee unity. The numerical values sugges that the first term represents the semi conductor junction and the second ter the other metallic parts of the diod For a pure metallic resistor the secon exponent should be unity, and in effect rather good matching can be obtained f exponent $m = 1$ if the other constants the equation are altered accordingly.

Equation 1 can easily be differentiate and integrated, and lends itself well f analytical work (see the appendix). Th additive terms facilitate separation of e

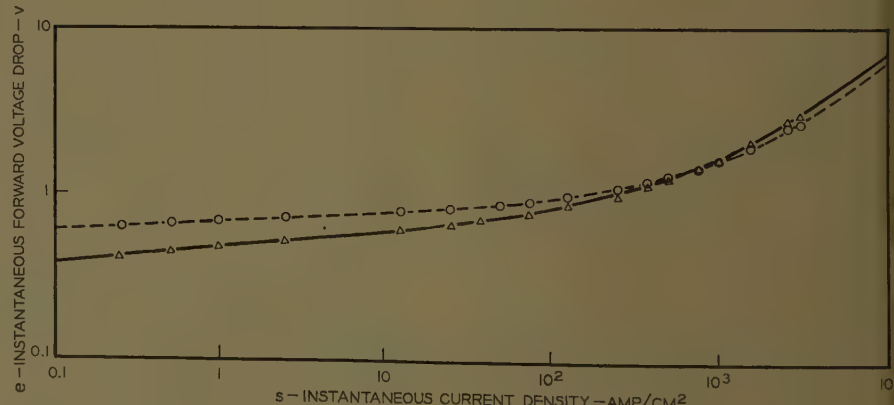


Fig. 3. Temperature influence on instantaneous forward voltage drop for alloyed junctio diodes

△ Solid line: $e = 0.49(s)^{0.0931} + 1.64(10^{-3})(s)^{0.8932}$, T = 423 K
○ Dashed line: $e = 0.70(s)^{0.0468} + 1.40(10^{-3})(s)^{0.8932}$, T = 298 K
F = 0.396 cm

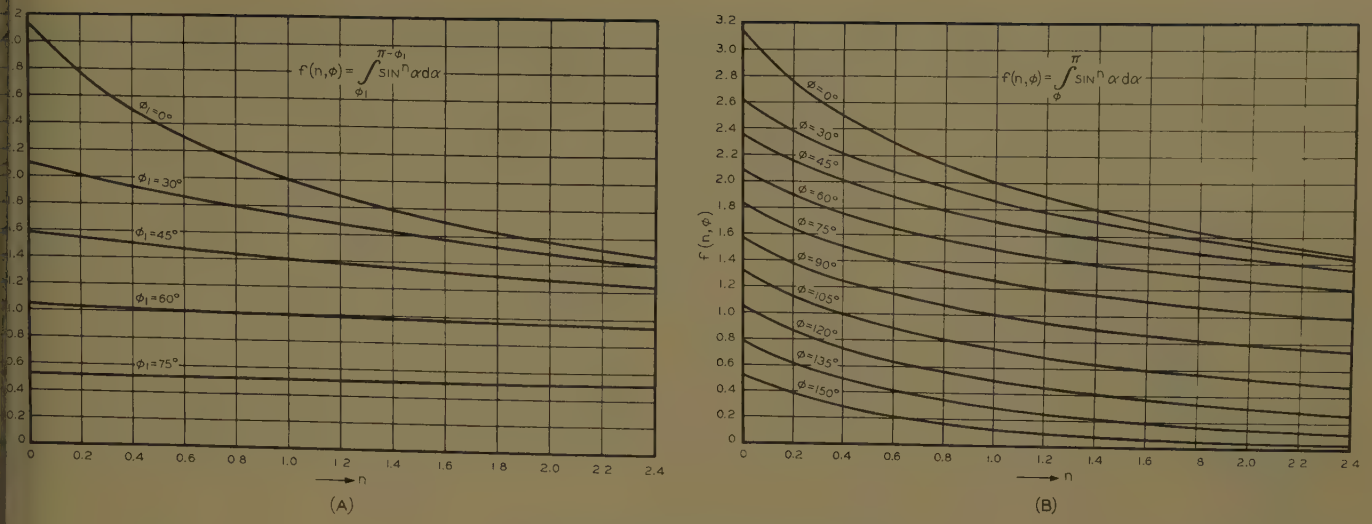


Fig. 4. Average forward voltage-drop integral $\sin^n \alpha d\alpha$

A— $\int_{\phi_1}^{\pi-\phi_1} \sin^n \alpha d\alpha$ B— $\int_{\phi_1}^{\pi} \sin^n \alpha d\alpha$

ects caused by the junction and by other arts of the diode.

Temperature Influence

To study the influence of the junction emperature, tests were made at 25 C and 50 C junction temperature with alloyed iodes having 0.396-cm² junction area. The result is shown in Fig. 3. The test oints follow the equations given in the igure. These two equations and the ne just given for the alloyed diodes can an be expressed in the general form

$$=(f_1/T)(s^{\vartheta_1 T}/s^{\vartheta_2})+h_1(1+\beta T)s^m \qquad (2)$$

Numerical values of the constants and exponents for the alloyed diodes are: $f_1=208$, $g_1=3.70\times10^{-4}$, $g_2=6.35\times10^{-2}$, $h_1=1.92\times10^{-6}$, $\beta=2.31\times10^{-3}$, $m=0.8932$.

Further test data seems to indicate that he temperature limits of validity for equation 2 are 25 C and 150 C, but that still rather good approximation can be obtained over the temperature range −50 C to 200 C by using a slightly lower value of f_1 than previously given.

As the second term in equation 2 is much smaller than the first term at a current density of 1 amp/cm², the forward voltage drop at this current density is nearly inversely proportional to the absolute temperature. This relationship holds approximately true between 0.1 and 10 amp/cm² within the temperature limits given. The influence of the temperature in the exponent of the first term is rather small, even at high current densities. Increasing the temperature from 25 to 150 C increases the factor $s^{\vartheta_1 T}$ about 10 amp/cm² and 40% at 1,000 amp/cm².

The second term includes a positive temperature coefficient as for a metallic resistor. The magnitude of the coefficient is, however, lower than for the metals composing the diode.

The result of the addition of both terms is that within a certain range of current density the forward voltage drop is nearly independent of temperature, and that there is a cross-over point as seen in Fig. 3.

Average Forward Voltage Drop

It is often desirable to determine the average forward voltage drop at a given junction temperature and for a specified conduction angle and average current. For sinusoidal current flow the average forward voltage drop over a full cycle is

$$\bar{E}_{\phi_1\phi_2}=\frac{1}{2\pi}\left[f(\hat{S})^{\vartheta}\int_{\phi_1}^{\phi_2}\sin^{\vartheta}\alpha d\alpha+ h(\hat{S})^m\int_{\phi_1}^{\phi_2}\sin^n\alpha d\alpha\right] \qquad (3)$$

The integral $\int_{\phi_1}^{\phi_2}\sin^n\alpha d\alpha$ has been determined for the two most usual cases; $\phi_2=\pi-\phi_1$, and $\phi_2=\pi$, and is represented in Figs. 4(A) and 4(B), for values of n from 0–2.4. The average forward voltage drop for various conduction angles corresponding to the instantaneous forward characteristic of the alloyed diodes from Fig. 1, but for 140 C junction temperature, is shown in Fig. 5. The junction temperature of 140 C was chosen, because it is a representative value for usual industrial applications

Forward Power Losses

The forward power loss at a constant junction temperaure is

$$\bar{P}_{\phi_1\phi_2}=\frac{F}{2\pi}\left[f(\hat{S})^{\vartheta+1}\int_{\phi_1}^{\phi_2}\sin^{\vartheta+1}\alpha d\alpha+ h(\hat{S})^{m+1}\int_{\phi_2}^{\phi_1}\sin^{m+1}\alpha d\alpha\right] \qquad (4)$$

In Fig. 6 the forward power loss is

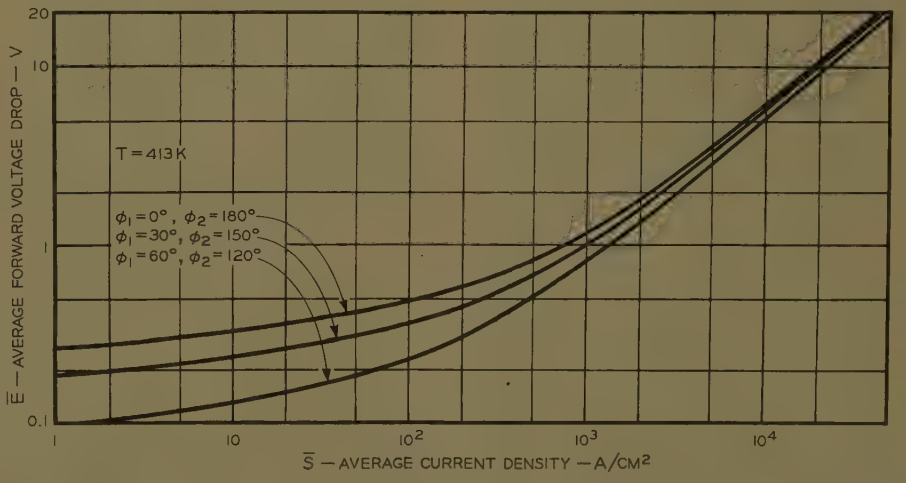


Fig. 5. Average forward voltage drop versus average current density for several conduction angles

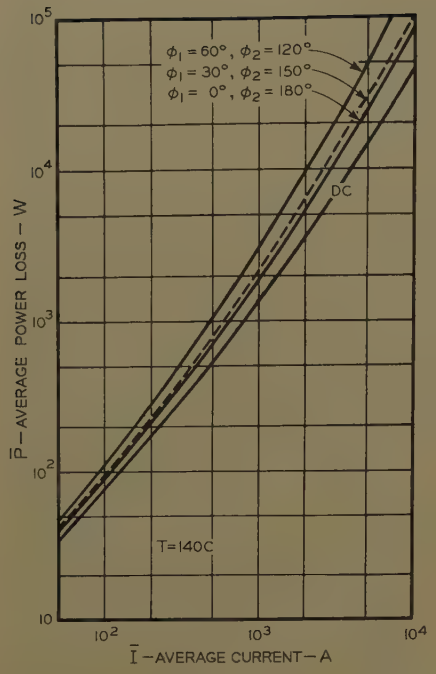


Fig. 6. Forward power loss versus average current at several conduction angles for a alloyed diode with 1.98-cm² junction area

shown for a diode having 1.98-cm² junction area and a junction temperature 140 . At a given average current the power loss for 120 degrees (deg) conduction (from 30–150 deg, is only approximately 10% higher than for 180-deg conduction, whereas for 60-deg conduction (from 60-120 deg) corresponding percentage is 30%. It should be observed that the power loss for a small conduction angle $\phi_2 - \phi_1$ varies considerably with ϕ_1. It makes a great difference if the conduction is from ϕ_1 to $\pi - \phi_1$, or from ϕ_1 to π, the losses in the latter case for the same average current density being much higher.

Frequently, the average forward voltage drop as function of average current has been determined for a diode by direct measurements, and it is desired to establish the power losses from these data. This can be done by introducing the form factor ψ from equation 13 (see the appendix). The average power loss for any conduction angle can then be expressed by

$$\bar{P}_{\phi_1\phi_2} = (\bar{E}\bar{I}\psi)_{\phi_1\phi_2} \qquad (5)$$

The form factor ψ is a weak function of current density and temperature, but varies strongly with the angles ϕ_1 and ϕ_2. Once the form factor has been determined for one type of diode, e.g., alloyed-junction diodes, it can be used for calculating the power losses in such diodes, even if the characteristics of the same vary considerably from sample to sample. For the alloyed-junction diodes investigated, ψ has been calculated at two temperatures, and for the three most common conduction angles. The result is shown in Fig. 7. For 3-phase and 6-phase operation ψ is nearly constant for current densities from 1 to 10^4 amp/cm², whereas for single-phase operation there is an increase in ψ of about 10% between 50 and 10^4 amp/cm². The form factor ψ for the diffused-junction diodes, with the forward characteristic shown in Fig. 2, has also been calculated. The average discrepancy between these values and the one shown in Fig. 7 for the alloyed-junction diodes is less than 1%, and the maximum deviation is less than 3%.

Pulse Load

For continuous operation an average current density of 130 amp/cm² (corresponding to a peak density of about 400 amp/cm²) is rarely exceeded in large area devices. However, under surge current conditions of short duration, a diode may be subject to current densities of the order of 10^4 amp/cm². To determine the momentary temperature rise of the junction under such conditions, it is necessary to know the power loss. Test data of the average forward voltage drop are not often available for such high-current densities. The equations previously given allow calculation of the power loss at any current density of interest. In order to give a clear picture of the behavior of the characteristics at high-current densities, all curves are shown for current densities up to 10^4 amp/cm².

Conclusions

For silicon diodes manufactured by t same process, but of different junctio area, the forward voltage drop at equ current densities and junction temperatu is very closely the same. The insta taneous forward voltage drop can be e pressed as a power function of curre density and the average forward volta drop for any conduction angle and ov a certain temperature range can be ca culated.

The power losses can either be dete mined from the instantaneous characte istics or from measured values of t average forward voltage drop and t current. In the latter case a form fact must be introduced, which varies on slightly with the forward characteristi and the junction temperature.

Appendix

Assume

$$s = \hat{S} \sin \alpha \qquad (6)$$

$$e = f(s)^q + h(s)^m \qquad (7)$$

Then by definition and equation 5

$$\bar{S}_{\phi_1\phi_2} = \frac{1}{2\pi}\int_{\phi_1}^{\phi_2} s\,d\alpha = \hat{S}\,(\cos\phi_1 - \cos\phi_2)/2\pi \qquad (8)$$

By definition

$$\bar{E}_{\phi_1\phi_2} = \frac{1}{2\pi}\int_{\phi_1}^{\phi_2} e\,d\alpha \qquad (9)$$

Substituting equations 6 and 7 in equation gives

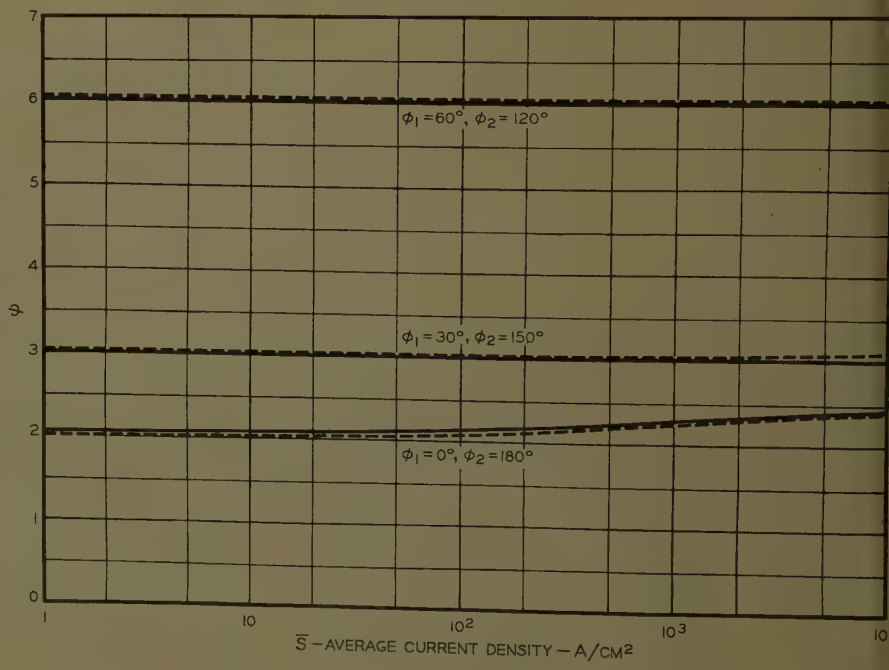


Fig. 7. Form factor ψ for alloyed-junction diodes at two junction temperatures

$$\phi_2=\frac{1}{2\pi}\left[f\left(\frac{2\pi\,\bar{S}_{\phi_1\phi_2}}{\cos\phi_1-\cos\phi_2}\right)^g\times\right.$$
$$\int_{\phi_1}^{\phi_2}\sin^g\alpha d\alpha+h\left(\frac{2\pi\,\bar{S}_{\phi_1\phi_2}}{\cos\phi_1-\cos\phi_2}\right)^m$$
$$\left.\int_{\phi_2}^{\phi_1}\sin^m\alpha d\alpha\right] \quad (9)$$

definition

$$\phi_2=\frac{F}{2\pi}\int_{\phi_1}^{\phi_2}es d\alpha \quad (10)$$

ostituting equations 5, 6, and 7 in equa- n 10 yields

$$\bar{P}_{\phi_1\phi_2}=\frac{F}{2\pi}\left[f\left(\frac{2\pi\bar{S}_{\phi_1\phi_2}}{\cos\phi_1-\cos\phi_2}\right)^{g+1}\times\right.$$
$$\int_{\phi_2}^{\phi_1}\sin^{g+1}\alpha d\alpha+h\left(\frac{2\pi\,\bar{S}_{\phi_1\phi_2}}{\cos\phi_1-\cos\phi_2}\right)^{m+1}$$
$$\left.\int_{\phi_2}^{\phi_1}\sin^{m+1}\alpha d\alpha\right] \quad (11)$$

Rearranging equation 11 gives

$$\bar{P}_{\phi_1\phi_2}=(\bar{E}\bar{S}\psi)_{\phi_1\phi_2}/F \quad (12)$$

Where

$$\psi_{\phi_1\phi_2}=\frac{2\pi}{\cos\phi_1-\cos\phi_2}\times$$

$$\frac{\int_{\phi_1}^{\phi_2}\sin^{g+1}\alpha d\alpha+\frac{h}{f}(\bar{S})^{m-g}\int_{\phi_1}^{\phi_2}\sin^{m+1}\alpha d\alpha}{\int_{\phi_1}^{\phi_2}\sin^g\alpha d\alpha+\frac{h}{f}(\bar{S})^{m-g}\int_{\phi_1}^{\phi_2}\sin^m\alpha d\alpha} \quad (13)$$

References

1. Temperature Rise of Solid Junctions Under Pulse Load, E. J. Diebold. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 593–98.

2. Properties of Silicon Power Rectifiers, E. F. Losco. *Ibid.*, vol. 74, Mar. 1955, pp. 106–11.

Discussion

A. Halacsy (Federal Pacific Electric Company, Newark, N. J.): Equations 1 and 2 seem to be but expression of experimental results in the languages of mathematics. Much insight could be gained into the operation of semiconductors if these equations would have some theoretical foundation. Did the author try to connect experimental results and semiconductor theory?

W. Luft: Real semiconductor diodes do not consist of a junction only, and do not have the forward characteristics predicted by semiconductor theory except at very low current densities. As silicon diodes generally are operated at high current densities, attempts to found empirical data for operating current densities on present theory are futile. The deviation from present theory is not due to other metallic parts composing a real device alone, as the resistance of such parts amounts to only about 5% of the total resistance of the semiconductor at the current densities in question.

On a Systematic Approximation to the Partition Method for Analysis of a Class of Nonlinear Systems

Y. H. KU FELLOW AIEE **A. A. WOLF** MEMBER AIEE **J. H. DIETZ** NONMEMBER AEE

THE DYNAMIC behavior of a broad class of physical systems can be described by ordinary nonlinear integro-differential equations of the type

$$(D)x(t)+F[t, x(t), x'(t),\ldots]=g(t) \quad (1)$$

with a specified set of initial conditions. (D) is a linear operator of the form

$$(D)=\sum_{r=-\lambda}^{\mu}f_rD^r,\ D^r=\frac{d^r}{dt^r} \quad (2)$$

in which the f_r are real constants restricted in such a way that $y(t)$, the impulsive response of the linear terms with constant coefficients defined in equation 3, is a function of exponential type and order one.

$$(t)=\frac{1}{2\pi j}\int_{Br}\frac{e^{st}}{Z(s)}\,ds \quad (3)$$

In equation 3 s is a complex variable $\sigma+j\omega$ and Br signifies an appropriate Bromwich contour enclosing the zeros of $Z(s)$.

$F[t, x(t), x'(t), \ldots]$ is a continuous, single-valued function expressible as a polynomial in t, $x(t)$, and/or integrals and derivatives of $x(t)$. It may contain nonlinear terms with constant coefficients, and linear and nonlinear terms with varying coefficients. The highest derivative of $x(t)$ appearing in the function is $x^{(\mu-1)}(t)$, where, as indicated in equation 2, μ is the order of the highest derivative of $x(t)$ in the linear terms $Z(D)x(t)$. The function satisfies the condition

$$\left|\frac{\partial^k F}{\partial[x^{(i)}(t)]^k}\right|<M_{ik}\epsilon^{-\alpha_{ik}t},\ t>0 \quad (4)$$

where M_{ik} and α_{ik} are positive real constants; $i=\ldots, -2, -1, 0, 1, \ldots, \mu-1$; and $k=1, 2, \ldots$

The forcing function, $g(t)$, is an analytic function of exponential type and order one for $t>0$.

Wolf[1,2] has shown that equation 1 as restricted possesses a unique solution and he has devised a general partition method of analysis which yields the exact solution as a power series in t with recursively related coefficients. Ku, Wolf, and Dietz have formalized this method in a transformation.[3] (In general, the exact solution of equation 1 is the linear combination of all the moments of the folded impulse response of the linear part.[1,2] Therefore the form of solution is a function of the partition point.)

Owing to the series form of the solution and to the multifold convolution sums arising from the polynomial representation of the nonlinearity, the exact partition method may be laborious to apply to the calculation of numerical solutions. In this paper the partition concept is extended to the development of a systematic method for obtaining approximate numerical solutions. The computational labor depends on the desired accuracy;

Paper 60-111, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 19, 1959; made available for printing December 9, 1959.

Y. H. Ku and J. H. Dietz are with The Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., and A. A. Wolf is with the Stromberg-Carlson Company, a division of the General Dynamics Corporation, Rochester, N. Y.

The authors are grateful to Dr. J. G. Brainerd, Director, The Moore School of Electrical Engineering, University of Pennsylvania, for his encouragement and support; to Thomas D. Truitt, of Electronic Associates, Inc., for his helpful criticism and suggestions; and to the National Science Foundation for their support. J. H. Dietz wishes to thank Dr. Douglas E. Mode, of The Moore School, for allowing him time to participate in this study.

it is usually less, however, than the labor required in the exact method.

After the nonlinear integro-differential equation is partitioned, it is converted to an integral equation of the convolution type, and approximate integration is employed to allow "integrating ahead" to the next discrete value of the solution. Since the method is numerical the nonlinear terms give rise to numerical values rather than to multifold convolution sums involving the coefficients of a power series. Similar methods based on the approximate evaluation of a convolution integral have been developed by Naumov[4] in the Soviet Union and Stout[5] in the United States. The approximate partition scheme is briefly compared with these methods.

Expressions are derived for estimating the error in the approximate solution. From these expressions the required value of the integration interval may be calculated for a predetermined accuracy.

Development of Method

The terms $F[t, x(t), x'(t), \ldots]$ are transferred to the right member of equation 1, and the difference between the actual forcing function and these transferred terms defines the auxiliary forcing function $a(t)$:

$$a(t)=g(t)-F[t, x(t), x'(t),\ldots] \tag{5}$$

(Some or all of the linear terms with constant coefficients except the term involving the highest derivative may also be transferred to the right side and included in the auxiliary forcing function.) Equation 1 then becomes

$$Z(D)x(t)=a(t) \tag{6}$$

which, when $a(t)$ is known, is a linear integro-differential equation with constant coefficients. The solution of equations of this type is the sum of a complementary function and a particular solution. The complementary function $b(t)$ is the solution of the homogeneous linear integro-differential equation with constant coefficients

$$Z(D)b(t)=0 \tag{7}$$

with the same initial conditions as equation 1. It is of the form

$$b(t)=\sum_{r=-\lambda}^{\mu-1} x^{(r)}(0+)\phi_r(t) \tag{8}$$

As restricted above, $g(t)$ and $F[t, x(t), x'(t), \ldots]$, and hence $a(t)$, are analytic functions of t for $t>0$. Since $y(t)$ is also analytic, the particular solution[6] of equation 6 is

$$x_p(t)=\int_{0+}^{t} a(\tau)y(t-\tau)d\tau \tag{9}$$

Then

$$x(t)=b(t)+\int_{0+}^{t} a(\tau)y(t-\tau)d\tau \tag{10}$$

The integrand of equation 10 is analytic with respect to both t and τ, and differentiation may therefore be carried out under the integral sign. Thus

$$x'(t)=b'(t)+a(t)y(0+)+\int_{0+}^{t} a(\tau)y'(t-\tau)d\tau \tag{11}$$

Since $y(t)$ is the solution of the linear integro-differential equation

$$Z(D)y(t)=0 \tag{12}$$

with initial conditions $y^{(i)}(0+)=0$ for $i<\mu-1$, and $y^{(\mu-1)}(0+)=1$; it follows that the second term in the right member of equation 11 is zero and, more generally, that

$$x^{(i)}(t)=b^{(i)}(t)+\int_{0+}^{t} a(\tau)y^{(i)}(t-\tau)d\tau \tag{13}$$

for $i<\mu$.

Approximate Solution

The solution of equation 1 by the general partition method consists of the simultaneous solution of equations 5 and 13 for the auxiliary forcing function, $a(t)$, and the substitution of this function in equation 10 to obtain $x(t)$. For the approximate solution the interval of t in which $x(t)$ is desired is divided into contiguous disjoint subintervals of equal length h; and the approximate values of $x(t)$, $x'(t)$, $g(t)$, $a(t)$, etc., for $t=t_k=kh$ are denoted by x_k, x_k', g_k, a_k, etc. Equation 5 then becomes

$$a_k=g_k-F(t_k, x_k, x_k'\ldots) \tag{14}$$

If the values of $x(t)$ and of the derivatives and integrals of $x(t)$ appearing in equation 1 are known at $t=t_{k-M}$, these values may be considered the initial conditions of equation 1. Then, for $t=t_k$, equation 13 becomes

$$x^{(i)}(t_k)=b^{(i)}(t_k)+\int_{t_{k-M}}^{t_k} a(\tau)y^{(i)}(t_k-\tau)d\tau, \qquad k\geq M \tag{15}$$

in which

$$b^{(i)}(t_k)=\sum_{r=-\lambda}^{\mu-1} x^{(r)}(t_{k-M})\phi_r{}^{(i)}(t_k-t_{k-M}) \tag{16}$$

If the convolution integral in equation 15 is approximated by an M-point rule for numerical integration, e.g., the trapezoidal rule ($M=1$), Simpson's rule ($M=2$), the three-eighths rule of Cotes ($M=3$), Weddle's rule ($M=6$), equation 15 may be written approximately

$$x_k^{(i)}=b_k^{(i)}+h\sum_{m=0}^{M}\beta_m a_{k-M+m}y_{M-m}^{(i)}, \qquad k\geq M \tag{17}$$

where

$$b_k^{(i)}=\sum_{r=-\lambda}^{\mu-1} x_{k-M}^{(r)}\phi_r{}^{(i)}(Mh) \tag{18}$$

and the β_m are real constant determined by the particular integration rule employed. For the trapezoidal rule $\beta_0=\beta_1=1/2$; for Simpson's rule $\beta_0=1/3$, $\beta_1=4/3$ and $\beta_2=1/3$.

Equations 14 and 17 are solved simultaneously for the discrete values a_k and $x_k^{(i)}$. Since equation 17 is valid only for $k\geq M$, it is first necessary to determine the values $x_1^{(i)}, \ldots, x_{M-1}^{(i)}$ for $-\lambda\leq i<\mu$ by means of the starting procedure described below. These values and the initial conditions of equation 1 are then used to calculate $a_0, \ldots, a_{M-1}$. Substituting g_0, t_0, and the appropriate $x_0^{(i)}$ in equation 14 yields a_0. Similarly, substituting g_1, t_1, and the appropriate $x_1^{(i)}$ in equation 14 gives a_1 etc.

In the first step of the simultaneous solution of equations 14 and 17 the quantities $x_M^{(i)}$ and the resulting quantities a_M are determined. The initial values $x_0^{(i)}$ are substituted in equation 18, and $b_M^{(-\lambda)}, \ldots, b_M^{(\mu-1)}$ are calculated. These quantities, together with $a_0, \ldots, a_{M-1}$, are substituted in equation 17. From equation 17 $x_M^{(-\lambda)}, \ldots, x_M^{(\mu-1)}$ are determined, and the appropriate $x_M^{(i)}$ are substituted in equation 14, to give an equation in which a_M is the unknown. If $x_M^{(\mu-1)}$ does not appear nonlinearly in this equation, the equation is linear in a_M. On the other hand, if $x_M^{(\mu-1)}$ does occur nonlinearly in equation 14, a nonlinear equation in a_M results. The last term in the convolution sum in equation 17 is $\beta_M a_k y_0^{(i)}$, which becomes $\beta_M a_M y_0^{(i)}$ for $k=M$. From the definition of the impulsive response $y_0^{(i)}$ is zero for $i<\mu-1$ and unity for $i=\mu-1$. Hence, if $x_M^{(\mu-1)}$ does not occur nonlinearly in equation 14, a_M either does not appear or appears linearly in the right member of equation 14. If $x_M^{(\mu-1)}$ does occur nonlinearly in equation 14, so also does a_M.

The value a_M, determined in the first step, is needed in the second, in which the values $x_{M+1}^{(-\lambda)}, \ldots, x_{M+1}^{(\mu-1)}$ and the resulting value a_{M+1} are obtained. The previously calculated values $x_1^{(-\lambda)}, \ldots, x_1^{(\mu-1)}$ are inserted in equation 18 as the new initial conditions. The resulting $b_{M+1}^{(-\lambda)}, \ldots, b_{M+1}^{(\mu-1)}$ and $a_1, \ldots,$

substituted in equation 17, to give
values $x_{M+1}{}^{(-\lambda)}, \ldots, x_{M+1}{}^{(\mu-1)}$. The
propriate $x_{M+1}{}^{(i)}$ are substituted in
nation 14 to obtain a_{M+1}. In the third
p the initial conditions are $x_2{}^{(-\lambda)}, \ldots,$
$^{\mu-1)}$, and the values computed are
$_{+2}{}^{(-\lambda)}, \ldots, x_{M+2}{}^{(\mu-1)}$ and the resulting
lue a_{M+2}. Continuing the procedure
lds the approximate discrete values of
solution and of the auxiliary forcing
nction.

Mathematically Specified Forcing Function

The method described above is appli-
ble when $g(t)$ is specified graphically,
pularly, or by an explicit mathematical
nction. For $g(t)$ given by an explicit
nction, however, the accuracy of the
ution can be improved by using the
act response of the linear terms with
nstant coefficients to the forcing func-
n. Substituting equation 5 in equa-
n 15, and defining, yields

$$(t_k) = \int_{t_{k-M}}^{t_k} g(\tau) y^{(i)}(t_k - \tau) d\tau \qquad (19)$$

ds to

$$(t_k) = b^{(i)}(t_k) + c^{(i)}(t_k) - \int_{t_{k-M}}^{t_k} F[\tau, x(\tau), x'(\tau), \ldots] y^{(i)}(t_k - \tau) d\tau \qquad (20)$$

F_k is defined by

$$= F(t_k, x_k, x_k', \ldots) \qquad (21)$$

d if, as before, the convolution integral
approximated by an M-point integra-
n rule, equation 20 becomes

$$^{(i)} = b_k{}^{(i)} + c_k{}^{(i)} - h \sum_{m=0}^{M} \beta_m F_{k-M+m} y_{M-m}{}^{(i)}, \quad k \geq M \qquad (22)$$

quations 21 and 22 are solved simultane-
sly for the quantities F_k and $x_k{}^{(-\lambda)}, \ldots,$
$^{(\mu-1)}$ in the same way that equations 14
d 17 are solved for a_k and the discrete
lues of $x(t)$ and its derivatives and
tegrals.

Since $g(t)$ and $y(t)$ are functions of
ponential type and order one, they
ssess Laplace transforms. Hence the
aplace transform of equation 19 is

$$c^{(i)}(t)] = \mathcal{L}[g(t)u(t-t_{k-M})]\mathcal{L}[y^{(i)}(t)] \qquad (23)$$

which

$$t)u(t-t_{k-M}) = \begin{cases} g(t), & t \geq t_{k-M} \\ 0, & t < t_{k-M} \end{cases} \qquad (24)$$

nce

$$g(t)u(t-t_{k-M})] = \mathcal{L}[g(t+t_{k-M})]\epsilon^{-st_{k-M}} \qquad (25)$$

d

$$\mathcal{L}[y^{(i)}(t)] = \frac{s^i}{Z(s)}, \quad i < \mu \qquad (26)$$

then equation 23 becomes

$$\mathcal{L}[c^{(i)}(t)] = \frac{s^i \mathcal{L}[g(t+t_{k-M})]}{Z(s)} \epsilon^{-st_{k-M}} \qquad (27)$$

Taking the inverse Laplace transforms of equation 27 for all $-\lambda \leq i < \mu$ and evaluating them at $t = t_k$ yields the quantities $c^{(i)}(t_k)$. This procedure is often less laborious than the direct evaluation of equation 19.

Elimination of Extraneous Solutions

It is observed above that if $x_k{}^{(\mu-1)}$ appears nonlinearly in equation 14, a nonlinear equation in a_k results. There may therefore be more than one real value of a_k satisfying equation 14. Since equation 1 as restricted possesses a unique solution, the auxiliary forcing function must also be unique, and all but one of the real values of a_k are extraneous. It is sometimes possible to eliminate the extraneous values by considering the physics of the system being analyzed. It is always possible to resolve the ambiguity in a_k by differentiating equation 1 once with respect to t, to give

$$DZ(D)x(t) + DF[t, x(t), x'(t), \ldots] = Dg(t) \qquad (28)$$

which has the same solution as equation 1 when the additional initial condition $x^{(\mu)}(0+)$ is provided by solving equation 1 for the highest derivative after the given initial conditions have been substituted therein. The order of equation 28 is $\mu' = \mu + 1$, but the order of the highest derivative of $x(t)$ appearing nonlinearly is still $\mu - 1 = \mu' - 2$. Hence, when equation 28 is partitioned and solved, the approximate discrete values of the auxiliary forcing function are obtained from a linear equation. There are therefore no extraneous solutions for the quantities a_k.

Changing the Interval h

Each step in the approximate solution of equation 1 is really a new solution of the equation based on the M previous solutions. Hence, to change the interval h to a new value h', it is necessary to obtain M previous values of a_k and of the quantities $x_k{}^{(i)}$, the intervals between which are h'. To double h one takes every other value of the $2M$ immediately preceding values as the M previous values; to triple h, every third value of the $3M$ immediately preceding values, etc. It is necessary to have determined nM values of the solution (including the initial values) the intervals between which are h if one wishes to let $h' = nh$ in the next step.

To decrease h to $h' = h/n$, one must interpolate between the previous values to obtain M consecutive values, the intervals between which are h'. Newton's formula for backward interpolation is applicable since the values between which interpolation is carried out are equidistant. The number of values of the solution that must previously have been calculated before interpolation can be employed to reduce h depends on the number of backward differences required to give the desired accuracy.

Starting Procedure

The quantities $x_1{}^{(i)}, \ldots, x_{M-1}{}^{(i)}$ for all $-\lambda \leq i < \mu$ which are needed to start the approximate solution may be determined from the Maclaurin series expansion of the solution:

$$x(t) = \sum_{n=0}^{\infty} x^{(n)}(0+) \frac{t^n}{n!} \qquad (29)$$

The quantities $x^{(n)}(0+)$ can be obtained from equation 1 and its initial conditions. Substituting the given initial conditions in equation 1 leads to $x^{(\mu)}(0+)$. Differentiating equation 1 and substituting $x^{(\mu)}(0+)$ and the given initial conditions therein yields $x^{(\mu+1)}(0+)$. Differentiating equation 1 a second time, and substituting $x^{(\mu)}(0+)$, $x^{(\mu+1)}(0+)$, and the given initial conditions therein gives $x^{(\mu+2)}(0+)$. This procedure is continued until as many coefficients as necessary have been calculated. The number required depends on the rapidity of convergence of the series solution, the magnitude of $t_{M-1} = (M-1)h$, and the desired accuracy of $x_{M-1}{}^{(\mu-1)}$.

If the value of h selected for the approximate solution is such that many terms of the series are required to compute $x_{M-1}{}^{(\mu-1)}$ to the desired accuracy, a submultiple of h is used to determine the quantities $x_1{}^{(i)}, \ldots, x_{M-1}{}^{(i)}$. Then the approximate solutions for $x_M{}^{(i)}, x_{M+1}{}^{(i)}, \ldots$ are computed until the procedure for increasing the interval can be applied.

Accuracy of Approximate Solution

The errors in the quantities $x_k{}^{(i)}$, defined by

$$e_{i,k} = x_k{}^{(i)} - x^{(i)}(kh) \qquad (30)$$

are due to the truncation errors in the approximation of the convolution integrals in the kth step and to the "inherited" errors from similar approximations in the

preceding steps. If the truncation errors in the kth step are represented by $T_{i,k}$, the exact solution of equation 1 at $t=kh$ is given by

$$x^{(i)}(kh)=\sum_{r=-\lambda}^{\mu-1}x^{(r)}[(k-M)h]\phi_r{}^{(i)}(Mh)+h\sum_{m=0}^{M}\beta_m a[(k-M+m)h]y_{M-m}{}^{(i)}+T_{i,k} \quad (31)$$

when an M-point integration rule is employed to approximate the convolution integrals.

Substituting equations 17 and 31 in equation 30 leads to

$$e_{i,k}=\sum_{r=-\lambda}^{\mu-1}e_{r,k-M}\phi_r{}^{(i)}(Mh)+h\sum_{m=0}^{M}\beta_m\{a_{k-M+m}-a[(k-M+m)h]\}y_{M-m}{}^{(i)}-T_{i,k} \quad (32)$$

Since the errors in the discrete values of the auxiliary forcing function, represented by the terms $a_{k-M+m}-a[(k-M+m)h]$, are nonlinear functions of the errors $e_{i,k-M+m}$, equation 32 represents a system of $\mu+\lambda$ simultaneous nonlinear difference equations. These equations may be solved recursively when the $T_{i,k}$ and the boundary conditions are known, but are not readily solved analytically.

Approximate solutions can be found, however, if the difference equations are linearized, and from these solutions the asymptotic values for large k can be estimated. If the approximate discrete values of the auxiliary forcing function are expanded in a Taylor series about the exact values of $x(t)$ and of the derivatives and integrals of $x(t)$ appearing in the function, and if the higher-order terms are neglected.

$$a_k\approx a(kh)+\sum_j A_j e_{j,k} \quad (33)$$

in which

$$A_j=\left[\frac{\partial a(t)}{\partial x^{(j)}(t)}\right]_{t=kh} \quad (34)$$

Then

$$a_k-a(kh)\approx\sum_j A_j e_{j,k} \quad (35)$$

where j extends over the orders of all the derivatives and integrals of $x(t)$ contained in the auxiliary forcing function. The quantities A_j are in general functions of the $x^{(j)}(kh)$, but over a limited range of the expansion, constant values may be assigned. When this is done and equation 35 is substituted in equation 32 a system of $\mu+\lambda$ simultaneous linear difference equations is obtained:

$$e_{i,k}=\sum_{r=-\lambda}^{\mu-1}e_{r,k-M}\phi_r{}^{(i)}(Mh)+h\sum_{m=0}^{M}\beta_m\left(\sum_j A_j e_{j,k-M+m}\right)\times y_{M-m}{}^{(i)}-T_{i,k} \quad (36)$$

Although equation 36 appears formidable in general form, in the solution of actual problems it is usually not. If the trapezoidal rule is used $M=1$, and each of the equations represented by equation 36 is a first-order difference equation. For Simpson's rule $M=2$, and second-order difference equations result. Also, in many problems only $x(t)$ or one of the derivatives or integrals of $x(t)$ appears nonlinearly, so that only one A_j need be considered. The solution of equation 36 may be further simplified by assigning a constant value to each of the quantities $T_{i,k}$ for all k. This is justified because the truncation error may be monitored during the solution, as is shown below, and the interval h may be changed if the absolute value of the truncation error exceeds a predetermined value.

Equations 36 may be solved by means of generating functions,[7] which are defined by

$$E_i(w)=\sum_{k=0}^{\infty}e_{i,k}w^k \quad (37)$$

where w is a complex variable $u+jv$ with $|w|<1$. Multiplying each of the $\mu+\lambda$ equations represented by equation 36 by w^k and summing the resulting expressions over all $k\geq 0$ yields a system of simultaneous transformed equations in which the unknown quantities are the $E_i(w)$. Solving this system of equations leads to $\mu+\lambda$ expressions of the form

$$E_i(w)=\frac{G_i(w)}{H(w)} \quad (38)$$

in which the functions $G_i(w)$ are polynomials in w, and include the truncation errors and the boundary conditions of the difference equations. $H(w)$, the characteristic equation of the system, is a polynomial in w.

Expanding equation 38 in partial fractions gives

$$E_i(w)=-\frac{\rho_{i0}}{1-w}-\frac{\rho_{i1}}{w_1}\frac{1}{1-\frac{w}{w_1}}-\frac{\rho_{i2}}{w_2}\frac{1}{1-\frac{w}{w_2}} \quad (39)$$

where 1, w_1, w_2, etc., are the poles of $E_i(w)$ and ρ_{i0}, ρ_{i1}, ρ_{i2}, etc., are the respective residues at these poles. Expanding each of the partial fractions in a geometric series and combining the coefficients of like powers of w yields the right member of equation 37. The first partial fraction in equation 39 represents the steady state error and arises from the assignment of constant truncation errors. The remaining partial fractions represent the transient errors, which, with increasing k, may either decrease toward zero or increase without limit, depending on whether the poles are absolutely greater than or less than unity. As k becomes large, the pole (or poles) smallest in absolute value dominates the solution. If, for example, this is a real pole w_q, the asymptotic value of the error for k large is given by

$$e_{i,k}\sim-\rho_{iq}-\frac{\rho_{iq}}{(w_q)^{k+1}} \quad (40)$$

If $|w_q|>1$, for k large

$$e_{i,k}\sim-\rho_{i0} \quad (41)$$

Estimation of Truncation Errors

After Milne[8] two approximate solutions $x_k{}^{(i)}$ and $x_k{}^{(i)*}$, are obtained so that

$$KT_{i,k}=x_k{}^{(i)*}-x_k{}^{(i)} \quad (42)$$

in which K is a real constant. Expanding the integrand of the convolution integral in equation 15 by means of Newton's formula for forward interpolation gives

$$a(\tau)y^{(i)}(t_k-\tau)=f(\theta)=a_{k-M}y_M{}^{(i)}+\theta\Delta[a_{k-M}y_M{}^{(i)}]+\frac{\theta(\theta-1)}{2!}\Delta^2[a_{k-M}y_M{}^{(i)}]+\ldots+\frac{\theta(\theta-1)\ldots(\theta-n+1)}{n!}\Delta^n[a_{k-M}y_M{}^{(i)}]+\ldots \quad (43)$$

in which

$$\theta=\frac{\tau-t_{k-M}}{h} \quad (44)$$

and

$$\Delta^n[a_{k-M}y_M{}^{(i)}]=\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}\times a_{k-M+j}y_{M-j}{}^{(i)} \quad (45)$$

From equation 44, $d\tau=hd\theta$. Hence

$$\int_{t_{k-M}}^{t_k}a(\tau)y^{(i)}(t_k-\tau)d\tau=h\int_0^M f(\theta)d\theta \quad (46)$$

Substituting equations 43 and 45 in equation 46, evaluating the integral in the right member for $M=1$, and neglecting all terms containing differences above the second yields

$$\int_{t_{k-1}}^{t_k}a(\tau)y^{(i)}(t_k-\tau)d\tau=\frac{h}{2}[a_{k-1}y_1{}^{(i)}+a_k y_0{}^{(i)}]-\Delta^2[a_{k-1}y_1{}^{(i)}]\frac{h}{12}$$

which the right member contains the proximate integral by the trapezoidal e and the principal part of the truncan error. Now, equation 46 is evaluated $M=2$, again neglecting all terms con- ning differences above the second:

$$\int_{t_{k-1}}^{t_k} a(\tau)y^{(i)}(t_k-\tau)d\tau = 2ha_{k-1}y_1^{(i)} + \Delta^2[a_{k-2}y_2^{(i)}]\frac{h}{3} \quad (48)$$

If the third differences of the integrands equations 47 and 48 are negligibly all, the second differences in the two uations are approximately equal. The ncipal part of the truncation error in uation 48 is therefore four times the ncation error arising from the trape- dal rule.

Substituting equations 47 and 48 in uation 15, and defining

$$x_k^{(i)} = \sum_{r=-\lambda}^{\mu-1} x^{(r)}(t_{k-1})\phi_r^{(i)}(h) + \frac{h}{2}[a_{k-1}y_1^{(i)} + a_k y_0^{(i)}] \quad (49)$$

$$x_k^{(i)*} = \sum_{r=-\lambda}^{\mu-1} x^{(r)}(t_{k-2})\phi_r^{(i)}(2h) + 2ha_{k-1}y_1^{(i)} \quad (50)$$

d

$$T_{i,k} = -\Delta^2[a_{k-1}y_1^{(i)}]\frac{h}{12} \quad (51)$$

ads to

$$x_k^{(i)}(kh) = x_k^{(i)} + T_{i,k} = x_k^{(i)*} - 4T_{i,k},\ i<\mu-1 \quad (52)$$

hence

$$T_{i,k} = x_k^{(i)*} - x_k^{(i)} \quad (53)$$

By means of equation 53 the truncation ror may be monitored during the solu- on. If it exceeds a predetermined value, decrease in the interval is required to aintain the desired accuracy of solution. milarly, if the truncation error falls elow a predetermined value, an increase h is possible.

A similar expression for monitoring the uncation error when Simpson's rule used is obtained by evaluating the integral in the right member of equation 46 for $M=2$ and for $M=4$, neglecting all terms containing differences greater than the fourth. The truncation error for Simpson's rule is given by

$$T_{i,k} = -\Delta^4[a_{k-2}y_2^{(i)}]\frac{h}{90} \quad (54)$$

if the fifth and higher differences are negligible.

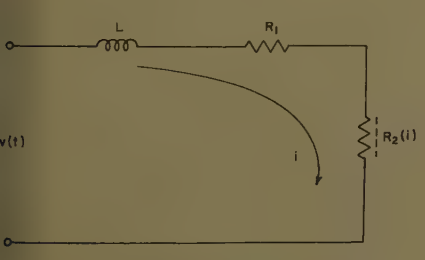


Fig. 1. First-order nonlinear circuit

Determination of Interval h

The error expressions developed above may be used to determine the interval h for the approximate integration. Suppose that $x(t)$ is desired in the interval $(0, t_{max})$ and that the absolute value of the error in the approximate solution must not exceed e_{max} in this interval. The maximum absolute value of the truncation error $T_{0,k}$ for each step in the solution must therefore be much less than e_{max}. Assigning a trial value of, say, 0.1 e_{max} to the truncation error, and noting that if the trapezoidal rule is used

$$T_{0,1} = -\Delta^2(a_0y_1)\frac{h}{12} \approx -\frac{d^2}{d\tau^2}[a(\tau\ y(h-\tau)]_{\tau=0+}\frac{h^3}{12} \quad (55)$$

one can compute a trial value of h. In evaluating the second derivative in equation 55, $a(0+)$, $a'(0+)$, and $a''(0+)$ are obtained from equation 1 and its initial conditions. For h small $y(0+)$, $y'(0+)$, and $y''(0+)$ are substituted for $y(h)$, $y'(h)$, and $y''(h)$. The value of h therefor obtained is substituted in equation 56

$$T_{i,1} \approx -\frac{d^2}{d\tau^2}[a(\tau)y^{(i)}(h-\tau)]_{\tau=0+}\frac{h^3}{12} \quad (56)$$

to compute the corresponding truncation errors for the derivatives and integrals of $x(t)$. These values, together with h, are substituted in equation 36, and the asymptotic solution for $e_{0,k}$ is obtained, with $k=t_{max}/h$ and the parameters A_j assigned values within their respective ranges so that $e_{0,k}$ is extremized. If the error obtained from equation 36 exceeds e_{max}, h must be decreased and the procedure repeated; if the error is much less than e_{max}, h is increased and the procedure repeated.

Equation 53 is used to monitor $T_{0,k}$ during the solution. If at a particular step a change in the interval h is indicated, the approximate error up to that step can be computed from the solution for $e_{0,k}$. The procedure for determining the initial value of h is then employed to compute the new value.

Application of Method

The series circuit shown in Fig. 1 contains a linear inductor L, a linear resistor R_1, and a nonlinear resistor R_2. At $t=0$ a voltage step of unit amplitude is applied. If $L=R_1=1$ and $R_2=i^2$, all in consistent units, and if there is no current flowing through L at $t=0$, the nonlinear differential equation for the current i is

$$\frac{di}{dt}+i+i^2=1 \quad (57)$$

with the initial condition $i(0+)=0$. To put equation 57 in the form of equation 1, write $i=x(t)$ and $di/dt=Dx(t)$. Then

$$(D+1)x(t)+x^2(t)=1 \quad (58)$$

with the initial condition $x(0+)=0$, from which

$$Z(D)=D+1 \quad (59)$$

$$F[x(t)]=x^2(t) \quad (60)$$

$$g(t)=1 \quad (61)$$

Partitioning equation 58 gives

$$(D+1)x(t)=a(t) \quad (62)$$

with

$$a(t)=1-x^2(t) \quad (63)$$

From equation 62 it is seen that

$$y(t)=\epsilon^{-t} \quad (64)$$

and

$$\phi_0(t)=\epsilon^{-t} \quad (65)$$

Let $h=0.2$ second. Then if the trapezoidal rule is used, equation 36 becomes

$$(1-0.1A_0)e_{0,k}-(0.819+0.082A_0)e_{0,k-1}=-T_0 \quad (66)$$

for which the solution is

$$e_{0,k}=-\frac{T_0}{0.181-0.182A_0}\times\left[1-\left(\frac{0.819+0.082A_0}{1-0.1A_0}\right)^{k-1}\right] \quad (67)$$

Now,

$$A_0=\frac{\partial a(t)}{\partial x(t)}=-2x(t) \quad (68)$$

Since $x(t)\geq 0$ from the physics of the problem, the transient error decreases toward zero with increasing k. The steady-state error is extremized when $A_0=0$. Hence for k large, $e_{0,k}\rightarrow -5.5T_0$. Equation 56 is used to approximate $T_{0,1}$. Since $a(0+)=1$ and $a'(0+)=a''(0+)=0$

$$T_{0,1}\approx y''(0.2)\frac{(0.2)^3}{12}=0.00055 \quad (69)$$

For large k, if $|T_0|\leq 0.00055$, $|e_{0,k}|\leq 0.003$.

Table I

t	x_k	x(t)
0	0.000	0.000
0.2	0.179	0.179
0.4	0.315	0.315
0.6	0.415	0.414
0.8	0.484	0.483
1.0	0.531	0.529
1.2	0.562	0.559
1.4	0.582	0.580
1.6	0.595	0.594
1.8	0.603	0.600
2.0	0.609	0.606

Substituting $g_0=1$ and $x_0=0$ in equation 14 gives $a_0=1$. Substituting $b_0=0$, $a_0=1$, $y_0=1$, and $y_1=0.819$ in equation 17 yields

$$x_1=(0.1)(0.819+a_1) \tag{70}$$

Combining equations 14 and 70 with $g_1=1$ leads to

$$a_1=1-(0.1a_1+0.082)^2 \tag{71}$$

for which there are two solutions: $a_1=-102.6$ and $a_1=0.967$. The ambiguity is resolved by noting that a positive voltage of unit amplitude is applied to the circuit, and the positive root is always used. Substituting $a_1=0.967$ in equation 70 gives $x_1=0.179$. Continuing the solution yields the values shown in Table I. Also tabulated for comparison is $x(t)$; $x(t)$ computed from the exact solution is

$$x(t)=-0.5+1.118\tanh 1.118(t+0.430) \tag{72}$$

The maximum difference between the approximate solution and the exact solution is 0.003, which occurs at $t=1.8$ and $t=2.0$ seconds. This agrees with the predicted error.

Second-Order Servo with Saturable Element

A servomechanism with a saturable element in the forward loop is shown in Fig. 2. The inner loop, containing only linear elements, consists of an actuator whose gain is K_a, a load with a moment of inertia J and a viscous-friction coefficient B, and a rate generator whose gain is K_b. The characteristic of the saturable element is

$$N(e)=K_c\tanh K_d e \tag{73}$$

in which K_c and K_d are constants determining the level at which saturation occurs and the small-signal gain, and e is the actuating error of the system, given by

$$e=r-c \tag{74}$$

where r is the reference input of the system, and c is the controlled variable.

For the inner loop

$$J\ddot{c}+(B+K_aK_b)\dot{c}=K_aN(e) \tag{75}$$

If $J=K_a=K_c=K_d=1$ and $B=K_b=0.5$, all in consistent units, eliminating c and its derivatives between equations 74 and 75 and combining the result with equation 73 gives

$$\ddot{e}+\dot{e}+\tanh e=\ddot{r}+\dot{r} \tag{76}$$

If $r(t)$ is a step of amplitude 10 applied at $t=0$, equation 76 becomes

$$\ddot{e}+\dot{e}+\tanh e=0,\ t>0 \tag{77}$$

with initial conditions $e(0+)=10$ and $\dot{e}(0+)=0$.

Letting $x(t)=e$, $Dx(t)=\dot{e}$, and $D^2x(t)=\ddot{e}$ puts equation 77 in the form of equation 1:

$$D(D+1)x(t)+\tanh x(t)=0 \tag{78}$$

with initial conditions $x(0+)=10$ and $x'(0+)=0$. Hence

$$Z(D)=D(D+1) \tag{79}$$

$$F[x(t)]=\tanh x(t) \tag{80}$$

$$g(t)=0 \tag{81}$$

Partitioning equation 78 gives

$$D(D+1)x(t)=a(t) \tag{82}$$

in which

$$a(t)=-\tanh x(t) \tag{83}$$

From equation 82 it is seen that

$$y(t)=1-\epsilon^{-t} \tag{84}$$

$$\phi_0(t)=1 \tag{85}$$

$$\phi_1(t)=1-\epsilon^{-t} \tag{86}$$

If the trapezoidal rule with $h=0.4$ second is used to approximate the convolution integrals, equation 36 becomes

$$e_{0,k}-(1+0.066A_0)e_{0,k-1}-0.330e_{1,k-1}=-T_0 \tag{87}$$

and

$$-0.2A_0e_{0,k}-0.134A_0e_{1,k-1}-0.670e_{1,k-1}=-T_1 \tag{88}$$

from the simultaneous solution of which the generating function for $e_{0,k}$ is

$$E_0(w)=\frac{(w-0.670w^2)T_0+0.330w^2T_1}{(w-1)[0.670w^2-(1.670+0.132A_0)w+1]} \tag{89}$$

Now,

$$A_0=\frac{\partial a(t)}{\partial x(t)}=-\operatorname{sech}^2 x(t) \tag{90}$$

the range of which is $(-1, 0)$. From equation 89 it is seen that $e_{0,k}$ is maximized for k large when $A_0=0$. For this value the partial-fraction expansion of equation 89 leads to the asymptotic solution

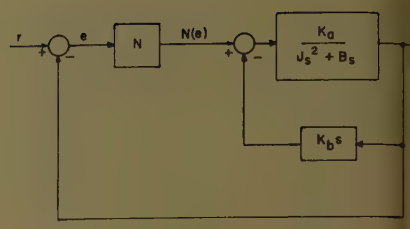


Fig. 2. Second-order nonlinear servo

$$e_{0,k}\sim -kT_0-(k-3)T_1 \tag{}$$

From equation 56, $T_{0,1}=-0.004$ $T_{1,1}=0.004$. Hence for $|T_0|\leq 0.004$ $|T_1|\leq 0.004$, $|e_{0,k}|\leq(2k-3)(0.004)$. $k=40$, $|e_{0,k}|\leq 0.31$.

Substituting $x_0=10$ in equation gives $a_0=-1$. For $x_0=10$ and $x_0'=$ equation 18 gives $b_1=10$. Substitut these two results in equation 17 for $M=$ and $i=0$ leads to $x_1=9.934$, for wh equation 14 yields $a_1=-1$. Final substituting $x_0'=0$, $a_0=-1$, and $a_1=$ in equation 17 for $i=1$ gives $x_1'=-0.3$ The values x_1 and x_1' are the new init conditions for the second step, wh yields $x_2=9.758$, $a_2=-1$, and x_2 -0.558. Continuing the solution le to the values shown in Table II a plotted in Fig. 3. Also plotted is system response determined by phase-plane method; see Fig. 4.[9] difference between the two responses

Table II

t	x
0	10
0.4	9
0.8	9
1.2	9
1.6	9
2.0	8
2.4	8
2.8	8
3.2	7
3.6	7
4.0	6
4.4	6
4.8	6
5.2	5
5.6	5
6.0	5
6.4	4
6.8	4
7.2	3
7.6	3
8.0	3
8.4	2
8.8	2
9.2	1
9.6	1
10.0	1
10.4	0
10.8	0
11.2	0
11.6	−0
12.0	−0
12.4	−0
12.8	−0
13.2	−0
13.6	−0
14.0	−0
14.4	−0
14.8	−0
15.2	0
15.6	0
16.0	0

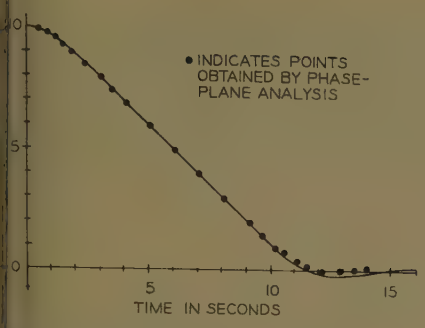


Fig. 3. Step response of nonlinear servo

10.4 seconds ($k=26$) is approxi-ately 0.2 which agrees with the esti-ated error.

omparison with the Methods of Naumov and Stout

Similar methods have been developed Naumov[4] in the Soviet Union and out[5] in the United States. Although rived in a different manner, these meth-s also consist in converting a non-ear differential equation to a nonlinear tegral equation of the convolution type. the notation used above it is of the rm

$$t)=c(t)+\int_0 F[t, x(t), x'(t), \ldots]y(t-\tau)d\tau \tag{92}$$

which $c(t)$ combines the response of e linear portion of the system to both e forcing function and the initial condi-ons, and then approximates the integral tween the limits $t=0$ and $t=kh$ by e trapezoidal rule. The resulting alge-aic equations are solved for the discrete lues x_k. If these equations are non-ear in x_k, they are solved graphically. xtraneous solutions are therefore pos-ble for certain nonlinearities, and no eans is provided for eliminating them. nce the convolution integral is approxi-ated over the interval $(0, kh)$ for each , it is necessary to compute all the lues $y(0+)$, $y(h)$, ..., $y(kh)$ and to nvolve this sequence with the sequence , x_1, ..., x_k in each step of the solution. ence the number of products of the type y_{k-m} that must be formed to deter-ine x_k is $1+2+\ldots+(k-1)+k$, or $^2+k)/2$, if $y(0+)=0$ and $(k^2+3k)/2$ if $0+)\pm 0$. When there are nonzero itial conditions it is necessary to com-te each of the k values of the $\mu+\lambda$ sponse functions due to the initial con-tions. If one of the derivatives of $x(t)$ pears nonlinearly and $x(t)$ is the de-ed solution, or if $x(t)$ and one of the rivatives of $x(t)$ both occur nonlinearly, e computational labor is doubled.

Another consequence of integrating from $t=0$ to $t=kh$ is that the interval h cannot be charged during the solution. Also, the more accurate integration rules, such as Simpson's rule, cannot be used. These rules require either an even or an odd number of intervals, and k takes on both even and odd values. Finally, estimating the error in the solution is difficult.

The approximate partition method differs from the methods of Naumov and Stout in that each step in the solution is a new solution of the nonlinear integro-differential equation with a new set of boundary conditions. If the trapezoidal rule is employed for the approximate evaluation of the convolution integral and if an additional approximate integration is carried out to allow the truncation error $T_{0,k}$ to be estimated, $k(\mu+\lambda+2)$ products of the form $a_{k+m-1}y_{1-m}$ are required to obtain all the values $x_k^{(-\lambda)}$, ..., x_k, ..., $x_k^{(\mu-1)}$. Only the values $y_0^{(i)}$ and $y_1^{(i)}$ for $-\lambda \leq i \leq \mu-1$ are needed. For Simpson's rule $k(2\mu+2\lambda+3)$ products of the form $a_{k+m-2}y_{2-m}$ must be formed. Since the value of each of the quantities $x_k^{(-\lambda)}$, ..., x_k, ..., $x_k^{(\mu-1)}$must be computed at each step af the approximate solution, as many as $k(\mu+\lambda)^2$ terms arising from the initial conditions for each step may have to be evaluated. Apart from the work connected with the actual evaluation of the nonlinear functions, the computational labor required in the approximate partition method is independent of the number of derivatives or integrals of $x(t)$ appearing nonlinearly.

Conclusions

A systematic method has been presented for extending the general partition method to the numerical analysis of a broad class of physical systems described by higher-order nonlinear integro-differential equations. Any M-point rule for the numerical integration of functions with equidistant ordinates may be used, the choice depending on the accuracy desired. The forcing function, nonlinearities, and varying parameters may be specified graphically, tabularly, or by explicit mathematical functions. A means for estimating the error in the approximate solution has been devised, together with a method for changing the interval h to maintain a predetermined accuracy.

Although developed for the analysis of systems described by a single nonlinear integro-differential equation, the method can be extended to the analysis of systems described by simultaneous nonlinear

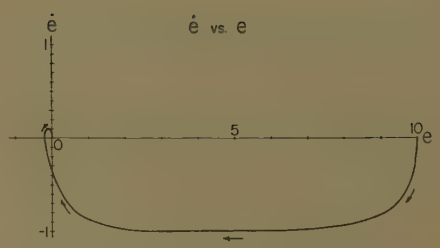


Fig. 4. Phase-plane trajectory for nonlinear servo

equations. A separate auxiliary forcing function can be defined for each equation and approximate discrete values of these functions solved for simultaneously. Matrix methods can be used after partitioning.

Since only M previously determined consecutive values of the solution are required for "integrating ahead" to the next value (and perhaps a few more if backward interpolation is used when the interval h is decreased during the solution), the method is suited for use with a general-purpose digital computer or with a digital differential analyzer.

The approximate partition method can be extended to the analysis of systems with stochastic forcing functions, just as the exact partition scheme is employed in the transform-ensemble method[10] for the analysis of such systems.

References

1. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems, A. A. Wolf. *Doctoral dissertation*, University of Pennsylvania, Philadelphia, Pa., 1958, chaps. II through IV.

2. Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems, Alfred A. Wolf. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 830–34.

3. Taylor-Cauchy Transforms for Analysis of a Class of Nonlinear Systems, Y. H. Ku, A. A. Wolf, J. H. Dietz. *National Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. III, "Circuit Theory," 1959, pp. 49–61; also *Proceedings*, Institute of Radio Engineers, vol. 48, no. 5, May 1960.

4. Eine Naeherungsmethode zur Berechnung der Uebergangsprozesse in Selbsttaetigen Regelungssystemen mit nichtlinearen Elementen, B. N. Naumov. Regelungstechnik: Moderne Theorien und ihre Verwendbarkeit, R. Oldenbourg, Munich, Germany, 1957, pp. 184–99.

5. A Step-by-Step Method for Transient Analysis of Feedback Systems with One Nonlinear Element, T. M. Stout. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 75, 1956 (Jan. 1957, section), pp. 378–90.

6. Elements of Pure and Applied Mathematics (book), Harry Lass. McGraw-Hill Book Company, Inc., New York, N. Y., 1957, pp. 195–96, 425–56.

7. An Introduction to Probability Theory and Its Applications (book), William Feller. John Wiley & Sons, Inc., New York, N. Y., vol. I, second edition, 1957, chap. XI.

8. Numerical Solution of Differential Equations (book), William E. Milne. John Wiley & Sons, Inc., 1953, chap. IV.

9. Analysis and Control of Nonlinear

SYSTEMS: NONLINEAR VIBRATIONS AND OSCILLATIONS IN PHYSICAL SYSTEMS (book), **Y. H. Ku.** The Ronald Press, New York, N. Y., 1958, pp. 14–16.

10. TRANSFORM-ENSEMBLE METHOD FOR THE ANALYSIS OF LINEAR AND NONLINEAR SYSTEMS WITH RANDOM INPUTS, **Y. H. Ku, A. A. Wolf.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 15, 1960.

11. ENGINEERING SOLUTIONS FOR CERTAIN NONLINEAR SYSTEMS, **J. H. Dietz.** *Master's thesis*, University of Pennsylvania, Philadelphia, Pa., 1959.

12. TRANSIENTS IN LINEAR SYSTEMS (book), **Murray F. Gardner, John L. Barnes.** John Wiley & Sons, Inc., 1942, chaps. V and VIII.

13. NUMERICAL MATHEMATICAL ANALYSIS (book), **James B. Scarborough.** The Johns Hopkins Press, Baltimore, Md., third edition, 1955, chaps. II, VII, VIII, and XI.

14. COMPLEX VARIABLE THEORY AND TRANSFORM CALCULUS (book), **N. W. McLachlan.** University Press, Cambridge, England, second edition, 1955 chap. III.

15. A NEW LINEAR OPERATIONAL CALCULUS, **Frank F. Bubb.** *Technical Report no. 6381*, U. S. Air Force, Dayton, Ohio, 1951.

16. NOTES ON TIME DOMAIN TECHNIQUES, **E. A. Guillemin.** *Lecture notes*, Massachusetts Institute of Technology, Cambridge, Mass., 1958.

17. HANDBOOK OF CALCULUS, DIFFERENCE AND DIFFERENTIAL EQUATIONS (book), **Edward J. Cogan, Robert Z. Norman.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1958, chap. XV.

Discussion

S. S. L. Chang (New York University, New York, N. Y.): I wish to congratulate the authors for their contribution to nonlinear system analysis. Their step-by-step method reduces the complexity of a nonlinear system calculation almost to that of numerical integration.

I should like to discuss two points:

1. Equation 2 defining the operator $Z(D)$ is an unnecessary restriction, which can be readily lifted with the state variable approach.

2. The partition into linear function $Z(D)x$, and nonlinear function F is not unique, and sometimes it is not desirable to partition according to the block diagram, or natural grouping of terms in the differential equation. A better way of doing this is perhaps to select the linear function $Z(D)x$ which best represents the entire system.

Consider a dynamical system described by the set of equations

$$\frac{dv_i(t)}{dt}=\sum_j q_{ij}v_j(t)+p_i a(t) \tag{93}$$

$$a(t)=g(t)-F[t, v_i(t)] \tag{94}$$

where $i, j=1, 2\ldots n$, and q_{ij} and p_i are constants. Equation 93 can be written as

$$sV_i(s)=\sum_j q_{ij}V_j(s)+v_i(0+)+p_iA(s) \tag{95}$$

Let the matrix $\Phi(s)$ be defined by

$$\bar{\Phi}(s)\equiv[s\bar{\mathrm{I}}-\bar{\mathrm{Q}}]^{-1} \tag{96}$$

where $\bar{\mathrm{I}}$ is the unit matrix, and $\bar{\mathrm{Q}}$ is the matrix with q_{ij} as its elements. Let $\Phi_{ij}(s)$ denote the elements of $\bar{\Phi}(s)$. Equation 95 can be written as

$$V_i(s)=\sum_j \Phi_{ij}(s)[v_j(0+)+p_jA(s)] \tag{97}$$

Its solution is

$$v_i(t)=\sum_{j=1}^{j=n} v_j(0+)\phi_{ij}(t)+\int_{0+}^{t} a(\tau)\sum_{j=1}^{j=n} p_j\phi_{ij}(t-\tau)d\tau \quad \text{for } i=1,2\ldots n \tag{98}$$

where $\phi_{ij}(t)$ is the inverse transform of $\Phi_{ij}(s)$. Equation 98 is exactly the same as equation 13 when the following correspondence is noted:

Ku-Wolf-Dietz Form	Generalized Form
$x^{(i)}(t)$	$v_i(t)$
$\phi_r{}^{(i)}(t)$	$\phi_{ij}(t)$
$y^{(i)}(t)$	$\sum_{j=1}^{j=n} p_j\phi_{ij}(t)$

Equations 14 through 18 can be readily rewritten into the generalized form. To see how the above format applies to systems with the fractional form of $Z(D)$, let

$$Z(D)=\frac{\sum_{r=0}^{r=n} a_rD^r}{\sum_{r=0}^{r=m} b_rD^r}, \quad n>m \tag{99}$$

Equations 17 and 18 in their original form no longer holds since $y^{(i)}(t)$ contains δ-functions, $i>n-m-1$. However, equation 1 can be rewritten into equation 98 by letting

$$v_r=x^{(r-1)} \quad \text{for } r=1, 2\ldots n-m$$

$$v_\lambda=\frac{dv_{\lambda-1}}{dt}-p_{\lambda-1}a(t) \text{ for } \lambda=n-m+1,\ldots n.$$

The matrix elements q_{ij} are given by

$$q_{ij}=1 \text{ for } j=i+1,\ i\neq n$$

$$=0 \quad j\neq i+1,\ i\neq n \tag{100}$$

$$q_{nj}=-a_{j-1}/a_n \tag{101}$$

The constants p_λ are determined by equating

$$\frac{1}{Z(s)}=\sum_j p_j\Phi_{ij}(s) \tag{102}$$

On both sides of equation 102, the denominator polynomials are automatically the same because of equations 100 and 101. By equating the $m+1$ coefficients of s in the numerator polynomials, the $m+1$ unknowns p_j with $j=n-m,\ldots n$ are determined.

After equation 1 is rewritten into equation 93, the generalized forms of equations 17 and 18 can be used to calculate $x(t)$ or other system variable, since the $\phi_{ij}(t)$'s do not involve δ-functions.

With regard to a better way of partition, we note that the method under discussion is equivalent to placing a sampler and multirate digital shaping filter after the nonlinear element. Since the digital shaping filter can never reproduce the original waveform, it appears that the less the signal is chopped up the better. The partition should be made such that F is the smallest possible. Consider the equation $0.01\ddot{x}+\sinh\dot{x}+x=$ Closer approximation is obtained by writing

$$Z(D)=0.01D^2+D+1$$

$$F=\sinh\dot{x}-\dot{x}$$

than using $0.01D^2+1$ as $Z(D)$ and $\sinh$ as F.

Y. H. Ku, A. A. Wolf, and **J. H. Dietz:** We wish to thank Prof. Chang for his thoughtful discussion and suggested generalization of the method. Taking up his two points in reverse order, the authors wish to make it clear that, according to the partition theorem the equation need not be partitioned according to the natural grouping of linear and nonlinear terms, and attention is called to the parenthetical note in the section "Development of Method." The solution is unique and independent of the way in which the equation is partitioned. In general the auxiliary forcing function should be defined in such a way that the truncation errors in the approximate convolution integrals are minimized.

In regard to Prof. Chang's first point it is noted that the only restriction on the operator $Z(D)$ is that its impulsive response given in equation 3, be of the exponential type and order one. For simultaneous integro-differential equations it is indicated in the conclusions that each equation in the set is properly partitioned. The resulting set is therefore

$$\sum_{k=1}^{n} Z_{jk}(D)x_k(t)=a_j(t) \tag{103}$$

with $j=1, 2,\ldots, n$, and

$$a_j(t)=g_j(t)-F_j[t, x_j(t),\ldots] \tag{104}$$

The solutions of 103 are of the form

$$x_k{}^{(i)}(t)=b_k{}^{(i)}(t)+\sum_{j=1}^{n}\int_{0+}^{t} a_j(\tau)y_{jk}{}^{(i)}(t-\tau)d\tau \tag{105}$$

where $k=1, 2,\ldots, n$, and the complementary functions $b_k(t)$ are obtained from the simultaneous solution of

$$\sum_{k=1}^{n} Z_{jk}(D)b_k(t)=0 \tag{106}$$

with the same initial conditions as equation 103.

In the particular integrals of equation 105 the $y_{jk}{}^{(i)}(t)$ are defined by

$$y_{jk}{}^{(i)}(t)=\frac{1}{2\pi j}\int_{Br}\frac{s^i\Delta_{jk}(s)}{\Delta(s)}\epsilon^{st}ds \tag{107}$$

where $\Delta(s)$ is the transformed system determinant for the linear terms with constant coefficients, the elements of which are $Z_{jk}(s)$, and $\Delta_{jk}(s)$ is the cofactor of $Z_{jk}(s)$. Equation 105 generalizes equation 13.

If

$$Z_{jk}(s)=s^{\mu_{jk}}+(\text{lower-order terms}) \tag{108}$$

$$\frac{s)}{)} = s^{-\nu jk} + (\text{higher-order terms}) \quad (109)$$

re $\nu_{jk} \geq \mu_{jk}$. Since in a physical system s greater than or equal to the number of gy-storage elements associated with , the order of the highest derivative of $x_k(t)$ that must be evaluated in the approximate partition method is less than or equal to $\nu_{jk}-1$. Hence the problem of δ-functions in the $y_{jk}^{(i)}(t)$ does not arise.

Equations 95, 96, and 97 are not necessarily valid since in a nonlinear system or a system with varying parameters the "state variables" and the auxiliary forcing function are not necessarily of exponential type and order one, even though the forcing function is such a function. Hence the state variables and the auxiliary forcing function may not possess Laplace transforms. The solution of the partitioned equation given in equation 13 is valid if the auxiliary forcing function is an analytic function of t. This latter condition is ensured by equations 2, 3, and 4.

The Measurement of Temperature in Welding Arcs

H. C. LUDWIG
NONMEMBER AIEE

opsis: The most successful technique in measurement of arc plasma temperatures been one based on the Saha theory which cribes the particle population of gases in number of thermally excited states. e presence and variation in these tes are detected by spectroscopic ans. The general method, first intro- ed by Fowler and Milne in the measure- nt of star temperatures, was adapted to n-made arcs by Larenz and Bartels and s been progressively improved upon by number of investigators. The method scribed in this paper is essentially that Larenz and Bartels except that im- oved instrumentation has been applied to l in the procurement of the spectroscopic ta.

KNOWLEDGE of the temperature distribution within the arc plasma is cessary for determining its energy ansport properties. These properties e interrelated to the irreversible proc- ses of radiation, convection, diffusion, d conduction which vary with the mperature and composition of the asma. Problems are being encountered arc welding research which require a ore thorough understanding of the ermal energy transport in the arc asma.

An established welding arc is a type of ermionic phenomenon. Upon initiation transforms from a short-lived low- mperature glow discharge to a discharge very high temperature. Its attained ermal state is controlled by a balance the simultaneous production and loss electrically conducting particles.

The temperatures attained in welding cs vary considerably according to the mposition of the plasma and the power put. The maximum temperature in w-current arcs <100 amperes, for ex- mple iron electrodes in air, is about 6,000 K (degrees Kelvin),[1] while in tungsten electrode arcs shielded by helium or argon the maximum temperature may be as high as 30,000 K. The highest temperatures realized thus far in welding arcs have been attained by making use of > 100-ampere current conducted through inert gas environments. Although these temperatures greatly exceed those required to melt the base metal, the welding speed, a function of the rate of heat transport to the welding zone, is increased due to a higher thermal energy state attained in the arc plasma. Thus the measurement of temperatures attained in welding arcs is desired for practical as well as scientific reasons.

Several basic findings, by Saha,[2,3] Boltzmann,[4] Compton,[5] and others[6–10] have preceded the development of techniques for measuring the extremely high temperatures of arcs and chemical flames. The temperature of a gas is represented by a Boltzmann distribution of the energy taken up by the particles present.[4] The energy in arcs is absorbed by the particles in translational, rotational, or vibrational states as energy of excitation, ionization, and radiation. It has been established by the control of the thermal behavior of electric arcs that temperature equilibium, i.e., equality in the temperature of ions, electrons, and neutral atoms, is closely approached or attained particularly at pressures of 1 atmosphere or higher.[5–10] It has also been found that the emission from the plasma of an electric arc is representative of the temperature of the gas provided no self-absorption of the radiation occurs. Saha[2,3] extended the law of mass action to include the dissociation of polyatomic gases; also, the excitation and ionization of all gases. It is from the Saha equation of thermal ionization that particle density of electrons, ions, and atoms can be calculated for any given pressure. It is upon the knowledge of the composition of the arc plasma that a basis for temperature measurements can be made. At a pressure of 1 atmosphere some 10^{17}–10^{18} particles per cm^3 (cubic centimeter) are present. These consist of electrons, ions, and atoms.

From the work of Fowler and Milne,[9,10] the variation in the intensity of thermally excited spectral emission lines was determined through astrophysical observations. Their findings led to the conception of temperature measurement techniques for man-made arcs based on several different types of spectroscopic observations. The method of temperature measurement to be described here is due essentially to Larenz[11] and Bartels.[12,13] It is based upon using intensities of spectral lines as thermometers. The selection of various spectral lines representing the quantized emission of the gas being thermally excited in the plasma is made according to the range of temperature being measured. Spectral emission due to excited atoms is generally selected for the lowest temperature found in arcs whereas emission due to singly and doubly ionized atoms is utilized for the higher temperatures. Certain improvements in technique and instrumentation have been made which eliminate the need for photographic records of the spectral radiation used in the prior method for the measurement of line intensities. The work described here involves the measurement of the temperature distribution in the tungsten-electrode argon-gas arc widely employed in the welding industry.

Composition of the Arc Plasma

The dependence of the degree of ionization upon temperature in a thermally

Paper 60-241, recommended by the AIEE Electric Welding Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted September 23, 1959; made available for printing December 10, 1959.

H. C. Ludwig is with the Westinghouse Electric Corporation, Pittsburgh, Pa.

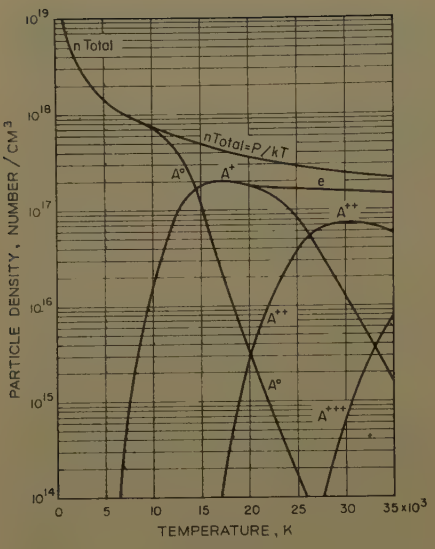


Fig. 1 (left). Argon-arc plasma composition at 1 atmosphere pressure

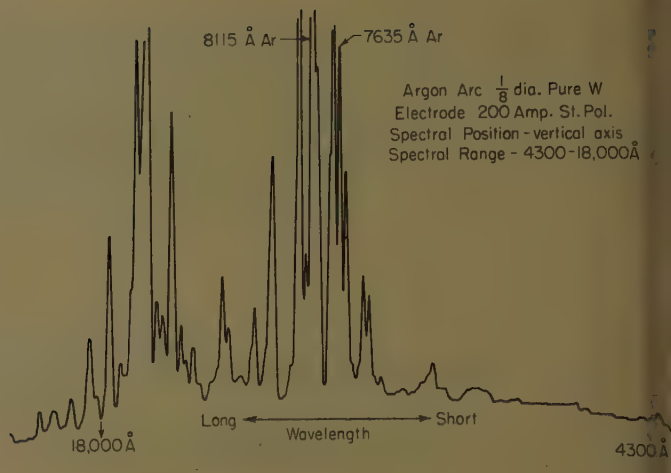


Fig. 2 (right). Spectrum of the tungsten, argon-shielded arc

excited gas can be determined according to the law of mass action,

$$\frac{n_+ n_e}{n_o} = f(T) \qquad (1)$$

where n_+, n_e, and n_o are particle densities of ions, electrons, and atoms, respectively. The ionization equilibrium as a function of temperature was developed by Saha according to the following equation:

$$\frac{n_+ n_e}{n_o} = \frac{U_+}{U_o} \frac{2(2\pi mkT)^{3/2}}{h^3} \exp(-E_i/kT) \qquad (2)$$

where

U_+ and U_o = the partition function for ions and atoms
m = mass of the electron
k = Boltzmann's constant
h = Planck's constant
E_i = energy of ionization or excitation

The particle densities of ions, electrons, and atoms may be determined through the simultaneous solution of equations 1, 2, and the following equations 3 and 4:

$$n_e = n_+ \qquad (3)$$

The total free charge in the case of thermal ionization is zero and

$$n_e + n_+ + n_o = n_{tot} \qquad (4)$$

where n_{tot} is the total particle density.

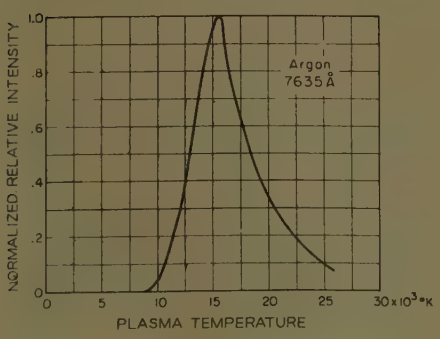


Fig. 3. Intensity as a function of temperature

The pressure as a function of temperature is determined from the ideal gas law,

$$P = (n_+ + n_e + n_o)kT \qquad (5)$$

The particle densities of ions, electrons, and atoms in thermally excited argon as a function of temperature[14] are shown in the semilogarithmic graph of Fig. 1. At any given temperature the composition of the gas is shown in kind and quantity.

Application of Emission Spectroscopy to Temperature Measurement

The emission from a stable electric arc represents a steady-state radiation of the transitions which the various excited and ionized particles undergo as the result of gain and loss of thermal energy. In returning to a state of lower energy than that attained, radiation is emitted which corresponds to the temperature of the particle. Thus, the emitted radiation, when properly resolved and measured, acts like a thermometer in measuring temperatures. Fowler and Milne, among others, derived the relationship between the intensity of a thermally excited spectral line and temperature. If the particle density of a certain excited level is given by Boltzmann's law and if the emission is from an optically thin layer, i.e., no self-absorption occurs,

$$I_{abs} = \frac{A}{4\pi} h\gamma \frac{g_a}{z_o} n_o \exp\left(-\frac{E_i}{kT}\right) \qquad (6)$$

where

I_{abs} = absolute intensity of light emitted in 1 cm³
A = transition probability of the energy to the ground state
g_a/z_o = partition function
γ = frequency of the emitted radiation
n_o = particle density of atoms in the ground state

For a given spectral line the variables in equation 5 are n_o and temperature (T), so

$$I_{rel} \sim n_o \exp\left(-\frac{E_i}{kT}\right) \qquad (7)$$

where

I_{rel} = relative intensity

By varying the temperature for a giv excitation potential E_i and using th corresponding neutral atom particl density calculated from the Saha equa tion 2, the proportionality shown in equa tion 7 can be determined. We now hav a relation for the spectral line intensit as a function of temperature correspond ing to a given excitation potential. Fo the determination of temperature in th tungsten, argon-shielded arc, the argo spectral line of 7,635 A (angstroms) wa chosen. This emission comes abou through an energy transition from 13.1 electron-volts to the ground state corre sponding to a wave number of 106, 237. cm^{-1}. The spectrum of the plasma of tungsten, argon-shielded arc shows line of neutral and singly ionized atoms. typical spectrum obtained using a record ing spectrometer is shown in Fig. 2. Th proportionality indicated in equation is shown graphically in Fig. 3 where tem perature is plotted on the abscissa an relative intensity normalized to unity fo the maximum value is shown on th

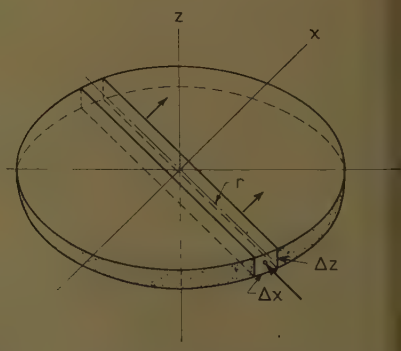


Fig. 4. Cross section through the arc plasm showing the orientation of a volume eleme

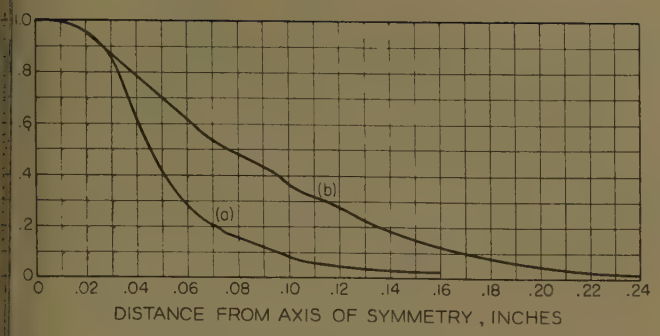


Fig. 5. Center-boundary variation of intensity argon 7,635 A

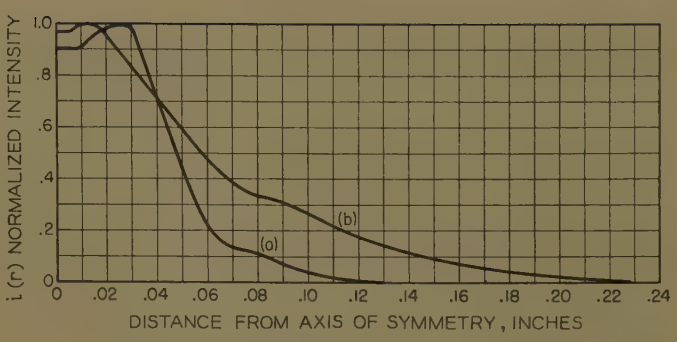


Fig. 6. Intensity as a function of arc radius

inate. In analyzing the curve of Fig. one can observe that in plotting ex- ential function exp $(-E_i/kT)$ against perature, unity is approached while the rticle density of neutral atoms decreases the temperature increases (Fig. 1). As esult, the curve of Fig. 3 has a maxi- m at a certain temperature. The sequence of the intensity of a spectral e having a maximum value correspond- to a certain temperature represents a ndard upon which the measurement extremely high temperatures can be sed.

gnificance of Spectral Line Intensity Measurements

The measurement of the spectral line tensities is obtained through the thick- ss of the arc plasma as it varies from the is radially outward to the boundary. ese data must be resolved so as to ob- in the local point intensity along the dius of the arc plasma. To obtain the cal point intensity as a function of the dius certain assumptions are made. It assumed that the plasma is symmetrical out its axis and that the temperature lls from a maximum value at the axis the arc boundary. The intensity of e spectral line as a function of the c radius, $i(r)$, is obtained from the ex- rimentally determined values of in- nsity, $i(x)$, through a 2-step mathema- tical process[15] of: (1) a differentiation of each experimental curve, and (2) an integration to be described. The experimental values, $i(x)$, assuming no absorption, are expressed as

$$i(x)=\sum_{-y_o}^{+y_o}[i(r)][\Delta y] \tag{8}$$

when considering, from Fig. 4, a section through the arc plasma where z is the vertical axis and $\Delta z \Delta x$ is the cross section of a small volume element. For symmetrical geometry

$$i(x)=2\int_{0}^{y_o} i(r)dy \tag{9}$$

If one replaces y with r as in independent variable, then

$$i(x)=2\int_{x}^{r_o}\frac{i(r)rdr}{\sqrt{r^2-x^2}} \tag{10}$$

The values $i(r)$ are determined by differentiating $i(x)$ and rearranging for the solution

$$i(r)=-\frac{1}{\pi}\int_{r}^{r_o}\frac{i(x)'dx}{\sqrt{r^2-x^2}} \tag{11}$$

Equation 11 is a so-called Abel integral equation. These mathematical processes must be carried out numerically. The use of a digital computing machine greatly facilitates the work.

The temperature as a function of the arc plasma radius, $T(r)$, i.e., the temperature distribution, is obtained from a cross-plot of the curve of the spectral line intensity as a function of the radius $i(r)$ and the line intensity as a function of temperature, $i(T)$, calculated from equation 7. It is necessary that a maximum occurs in the curve of line intensity as a function of the radius corresponding to the maximum value obtained in the curve of equation 7 as shown in Fig. 3. Thus, a fixed point is secured within range of the temperature distribution. This requirement is analogous to the condition of selecting a thermometer having a temperature range within the values to be measured.

Typical curves of $i(x)$ obtained from experimental measurements of a 300-ampere tungsten, argon-shielded arc are shown in the graph of Fig. 5. These curves were obtained from measurements made near the cathode (a) and near the anode (b) along planes perpendicular to the axis. Corresponding curves whose data were determined from equation 11 are shown in the graph of Fig. 6. The temperature distribution determined for the locations in the plasma described above are shown in the graph of Fig. 7. Isotemperature curves for the tungsten-electrode argon-shielded arc are shown in the graph of Fig. 8.

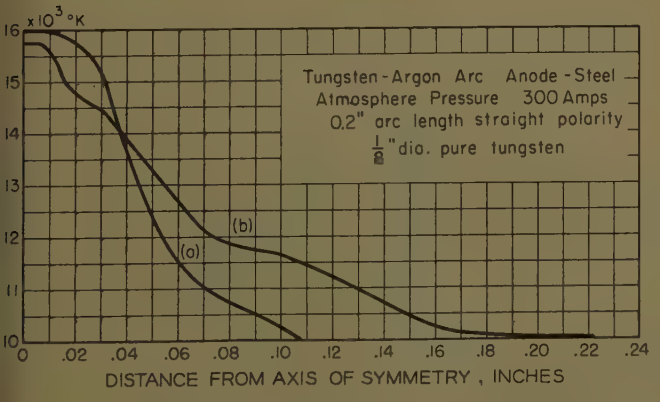


Fig. 7. Plasma temperature distribution

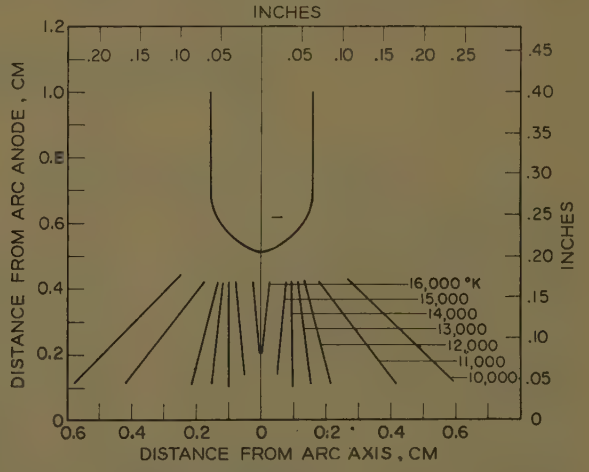


Fig. 8. Temperature distribution in an argon-tungsten arc, 300 amperes

The Technique and Experimental Procedure

In the experimental technique to be described, the spectral line intensity variation from the center to the boundary of the arc plasma is measured by using automatic recording devices instead of a photographic process. In addition to the need of darkroom facilities, photographic records of the spectra require corrections for nonlinear response of blackening to the variations in spectral line intensity.

The spectroscopic method described here makes use of a recording spectrometer of the monochromator type equipped with a light chopper and a quartz prism using single-pass optics. The receptor is a vacuum thermocouple. Provision has been made to use a photomultiplier tube to serve also as a receptor. The chopped signal is amplified as an alternating signal and converted to a unidirectional signal by synchronized circuit breakers. The amplified unidirectional signal is fed to a high-speed null-type potentiometer recorder. In obtaining the spectral line intensity variation from the center of the arc plasma to the boundary, the spectrometer is set to receive the wavelength of light corresponding to the chosen spectral line.

The image of the arc magnified several times is focused at the slit of the spectrometer and caused to scan across the slit at a constant rate. The record obtained on the high-speed recorder represents the spectral line intensity variation from one side of the arc to the other. The results obtained can be analyzed at once for geometric symmetry. Once the speed of the drive of the recorder, the rate for which the image of arc is scanned across the slit, and magnification of the arc are known, one can obtain the required lineal measurement of the distance from the center to the boundary of the arc. After normalization of the data obtained from the recorded curves, one proceeds as aforedescribed.

Summary

An improved technique for determining the temperature distribution in symmetrical and optically thin arc discharges has been developed based on a method described by Larenz and others. The improved technique makes use of instrumentation from which one obtains a direct record of the spectral line intensities. The need for a photographic record has been eliminated.

References

1. Arc Temperatures by an Optical Method, **C. G. Suits.** *Physics*, New York, N. Y., vol. 6, Oct. 1935, pp. 315–22.

2. Ionization in the Solar Chromosphere, **M. N. Saha.** *Philosophical Magazine*, London, England, series 6, vol. 40, 1920, pp. 472–881

3. **M. Saha.** *Zeitschrift für Physik*, Berlin, Germany, vol. 6, 1921, pp. 40–55.

4. Thermodynamics for Chemists (book), **S. Glasstone.** D. Van Nostrand Company, Inc., Princeton, N. J., 1947, pp. 101–02.

5. Theory of the Electric Arc, **K. T. Compton.** *Physical Review*, New York, N. Y., vol. 21, 1923, pp. 266–91.

6. Thermic Mechanism in the Column of the Arc Light, **L. S. Ornstein, H. Brinkman.** *Physica*, The Hague, Netherlands, vol. 1, July 1934, pp. 797–82.

7. **A. M. Kruithof.** *Ibid.*, vol. 10, 1943, pp. 493–501.

8. The Applicability of Saha's Formula to the Electric Arc, **A. M. Kruithof, J. A. Smi** *Ibid.*, vol. 11, 1944, pp. 129–43.

9. Intensities of Absorption Lines in Stella Spectra, **R. H. Fowler, E. A. Milne.** *Month Notices*, Royal Astronomical Society, London, En land, vol. 83, May 1923, pp. 403–24.

10. Statistical Mechanics (book), **R. H. Fowle** Cambridge University Press, Cambridge, Englan 1936.

11. **R. W. Larenz.** *Zeitschrift für Physik*, vol. 12 1951, pp. 327–64.

12. **H. Bartels.** *Ibid.*, vol. 127, 1950, p. 243.

13. **H. Bartels.** *Ibid.*, vol. 128, 1950, pp. 546–7

14. **G. Busz-Pueckert, W. Finkelnburg.** *Ibid* vol. 144, 1956, p. 244.

15. **H. Hormann.** *Ibid.*, vol. 97, 1935, pp. 539–6

Discussion

Thomas B. Correy (Hanford Atomic Prod ucts Operation, General Electric Company Richland, Wash.): Information in th literature on the electric arc phenomenon a applied to welding is very sketchy at bes This is a very good step in the direction explaining the mechanics of the electric ar in inert arc welding. The paper would b much more valuable if it treated the weld ing arc as a point-to-plane phenomenon fo both the straight and reverse polarity arc

H. C. Ludwig: I thank Mr. Correy for h discussion of this paper. It should be note that the work described was not intende to be a complete study of the temperatur distribution in the tungsten-argon arc but to describe a method by which tempera ture measurements on this type of arc coul be made. It is felt that the direct readin of spectral line radiation intensities is a improvement in the Larenz-Bartels Metho because it eliminates errors introduced b conversion of the radiation measurement from photographic film. Considerab time is also saved.

An Analysis of Transfer in Gas-Shielded Welding Arcs

W. J. GREENE
ASSOCIATE MEMBER AIEE

THE commercial success of a metal-arc-welding process depends in large measure upon the mode by which molten metal is transferred from the electrode to the workpiece. Excessively large drops will produce irregular beads which, in some cases, are bounded by regions of spatter. The presence of spatter not only mars the appearance of the joint but also represents a loss in deposition efficiency. This phase of the welding problem has occupied a prominent place in arc physics research during the past decade. Motion pictures showing transfer in slow motion is a familiar tool which has contributed to our knowledge of the various patterns and modes of metallic transfer with preset welding conditions.

At present, there is no formal theory for metallic transfer, although various aspects of of the problem have yielded t analysis by quantitative methods.[1] In contrast with previous investigation the subject will be approached on purel theoretical grounds to develop as much the basic theory as is consistent with th mathematical model assumed in the d velopment. In the interest of clarit the theory will be presented in three se tions. The first section will be limite to the formation and transfer of dro under the influence of gravitational forc and surface tension alone. These cond

Paper 60-240, recommended by the AIEE Electr Welding Committee and approved by the AIE Technical Operations Department for presentati at the AIEE Winter General Meeting, New Yo N. Y., January 31–February 5, 1960. Manuscri submitted April 8, 1959; made available for pri ing December 11, 1959.

W. J. Greene is with the Air Reduction Compan Inc., Murray Hill, N. J.

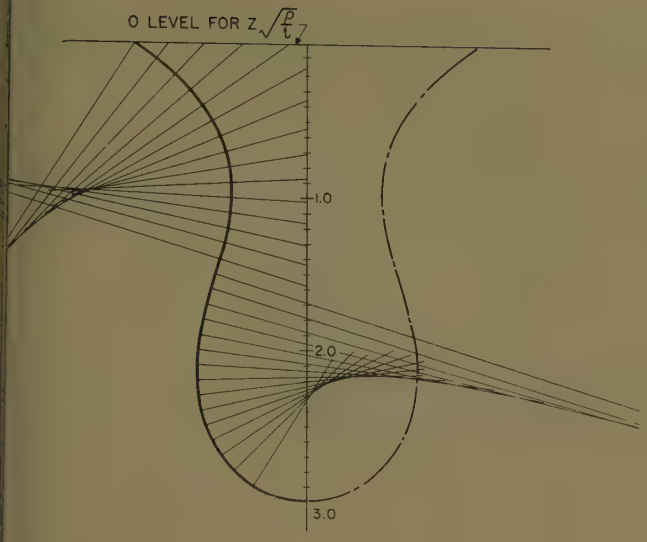


Fig. 1. Graphical integration to obtain drop file

...s are approximated in very-low-cur-...t arcs, where arc forces are negligible ...comparison with the other forces ...sent.

...n the second section, the formation ...d transfer of weightless drops will be ...sidered. In this case, arc forces and ...face tension are the only forces acting ...effect transfer.

...Finally, the third section will deal with ...three forces. In this fully developed ...m, it will be seen that terms can be ...lected into dimensionless groups not ...ike Reynolds, Prandtl, and Grashof ...mbers, which facilitate computation in ...at-transfer and fluid-flow problems. ...e of the dimensionless groups, called ... "transfer number," can serve as an ...lex to determine whether transfer in ... welding arc will be acceptable. ... means of this number, transition ...rents as well as maximum and zero re-...lsion currents can be located. With ... aid of an additional dimensionless ...up called the "drop size index," the ... of drops in the repulsion, pretransi-tion, and posttransition regions can be estimated.

Transfer Under Gravitational and Surface Tension Forces

Surfaces of revolution formed by surface tension at the boundary of a liquid under gravitational forces must satisfy the differential equation

$$\frac{d^2z}{drds}+\frac{1}{r}\frac{dz}{ds}=\frac{\rho z}{t} \qquad (1)$$

where

ρ = density
t = surface tension

A cylindrical co-ordinate system (r, θ, z) with the z-axis vertical is assumed. The increment ds is taken along the profile curve. Let

f = profile function = $r = f(z)$
f' = derivative of profile function = $dr/dz = f'(z)$
f'' = second derivative of profile function = $d^2r/dz^2 = f''(z)$

Eliminating ds from the equation 1:

$$-\frac{f''}{(1+f'^2)^{3/2}}+\frac{1}{f(1+f'^2)^{1/2}}=\frac{\rho z}{t} \qquad (2)$$

The solution of this equation under prescribed boundary conditions can best be obtained by graphical methods since the first two terms have geometric significance.

$$R_p=\frac{(1+f'^2)^{3/2}}{f''}$$

is the radius of curvature in the profile plane.

$$R_r=f(1+f'2)^{1/2}$$

is the radius of the inscribed sphere. Therefore,

$$\left\{\frac{1}{R_r}-\frac{1}{R_p}\right\}=\frac{\rho z}{t} \qquad (3)$$

where

R_r is limited to positive values
R_p is positive for concave outward curvature and negative for concave inward curvature

Terms in this equation may be made dimensionless to generalize graphical integration as follows:

$$\left\{\frac{1}{\sqrt{\frac{\rho}{t}}R_r}-\frac{1}{\sqrt{\frac{\rho}{t}}R_p}\right\}=\sqrt{\frac{\rho}{t}}\,z \qquad (4)$$

In the graphical integration process for drops supported from a vertical rod, the integration is started at the bottom of the drop, where

$$-\sqrt{\frac{\rho}{t}}R_p=\sqrt{\frac{\rho}{t}}R_r=\frac{2}{\sqrt{\frac{\rho}{t}}z}$$

The method used is similar to the circle of curvature method described by Lord Kelvin.[4] A typical integration is shown in Fig. 1. By performing a series

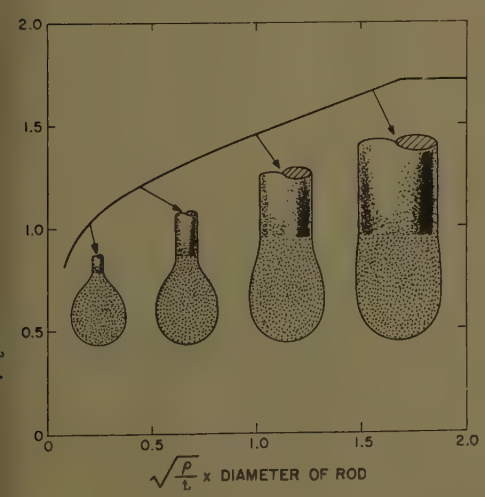


Fig. 2 (left). Effect of wire diameter on drop diameter

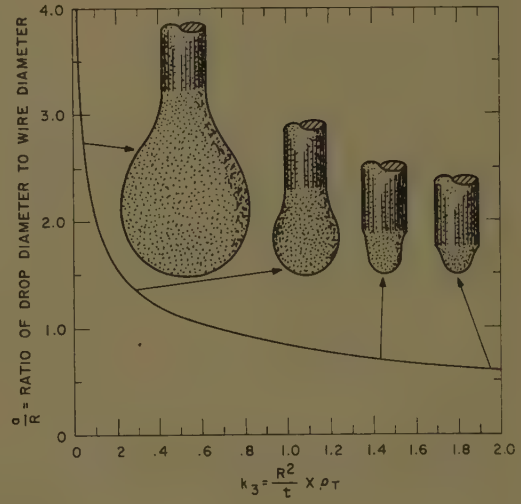


Fig. 3 (right). Effect of density on drop diameter

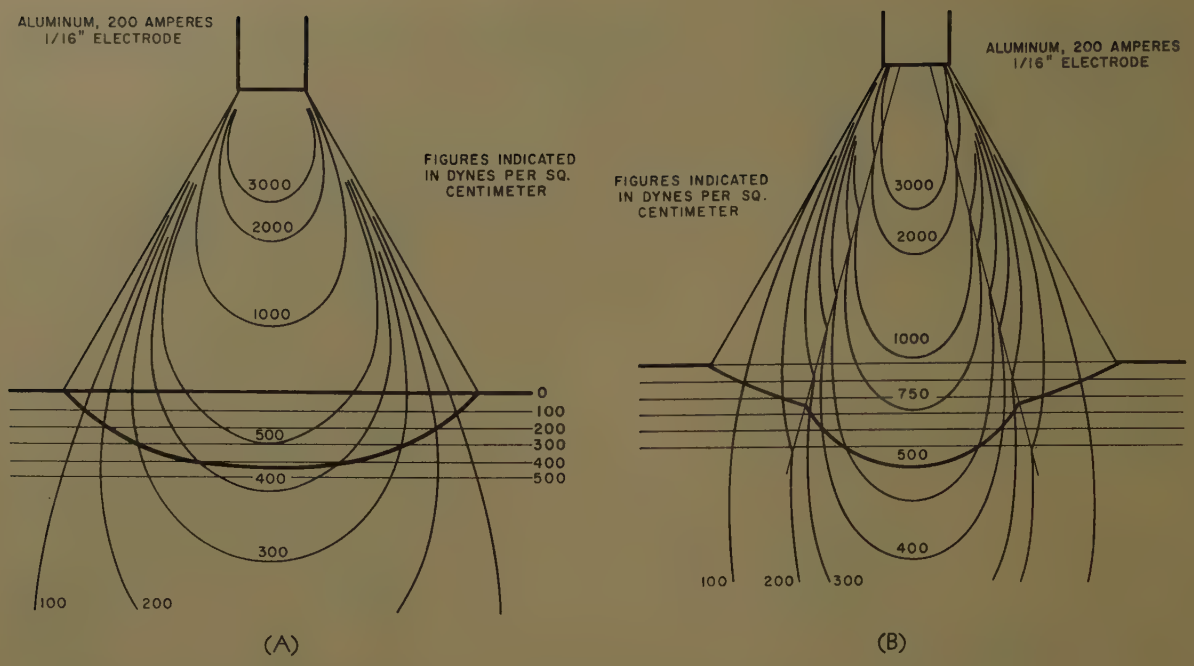


Fig. 4. Press distribution pool depres

A—Arc with central con

B—Arc central zon

of such integrations for various values of $\sqrt{\rho/t}z$ at the bottom of the drop, the relationship between the maximum diameter of the drop and the diameter of the neck can be obtained. This relationship is plotted in dimensionless form in Fig. 2. The largest drop which can be supported on the end of a rod is determined by setting the diameter of the rod equal to the diameter of the neck. The surface of the drop must terminate on the periphery of the rod if a stable solution exists. When the neck diameter is greater than rod diameter, no stable solution exists. The drop would then be transferred due to its own weight. If the neck diameter is less than the rod diameter, the drop could grow to a greater size before surface tension could no longer support it against gravitational forces.

At a critical lower limit of

$$\sqrt{\frac{\rho}{t}}\,z$$

at the base of the drop, the neck diameter and maximum drop diameter become equal. This diameter represents the maximum-size drops which can form. Drop diameter remains constant at this characteristic value for further increase in rod diameter. This occurs when $\sqrt{\rho/t}$ multiplied by rod diameter equals 1.7.

To show the effect of variations in the density of the weld metal, the data are replotted with the ratio of drop radius to electrode radius as ordinate and

$$\frac{R^2}{t}\,\rho_T = k_3$$

as abscissa (see Fig. 3), where R is the radius of the electrode wire. The maximum radius of drops which can be supported against gravity by surface tension steadily decreases with increased density of the weld metal.

The grouping of terms

$$\left(\frac{R^2}{t}\,\rho\right)$$

is dimensionless; and when calculated using any consistent set of units, the resultant number M will be termed the "drop size index."

When arc forces are not significant, the effective density of the weld metal ρ_T is equal to the actual density ρ due to gravity alone. Drop size can be determined for any electrode radius, surface tension, and density of weld metal by calculating the drop size index M and reading the corresponding ratio of drop radius to wire radius from Fig. 3.

Transfer of Weightless Drops Under the Action of Surface Tension and Arc Forces

As in the case of all fluid conductors, the plasma of the welding arc and the molten metal at the tip of the electrode are under slight positive pressure. This pressure originates from Lenz forces or motor action caused by the flow of electric current in the self-induced magnetic field.

By assuming that the plasma is conical in shape, forming an angle θ with the axis of the cone at the apex, the pressure at any point (r, ψ) is given by equation 5:

$$P = \frac{2\,I^2}{\pi r^2(1-\cos\theta)^2}\,\ln\frac{\cos\psi/2}{\cos\theta/2} \qquad (5)$$

where

P = pressure, dynes/cm² (per square cm meter)
I = current, abamperes
r = distance of point from apex measured cm
ψ = angle formed by r with the axis of cone
ln = natural logarithm (base = 2.71828)

The pressure pattern for a 200-amp arc with current distributed in a cone apex angle of rotation θ equal to degrees and arc length of 1 cm is shown Fig. 4(A). The depression of cathode pool resulting from this press is determined using the density aluminum. The apex of the cone intersected by an electrode wire assum to be 1/16 inch in diameter.

In spray-type arcs containing a w defined inner cone of metallic vapor, current flow may favor the more ioni path afforded by the metallic vapor. 4(B) is a plot of the pressure pattern a cathode pool depression, assuming t half the current (100 amperes) is distr uted evenly over the entire cone and other half is distributed evenly over inner cone which forms an angle of ro tion of 15 degrees at the apex.

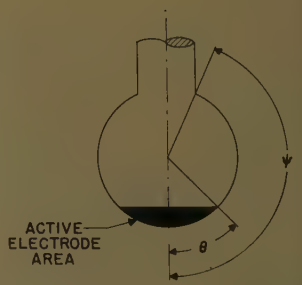


Fig. 5. Central angles for determinatio arc forces

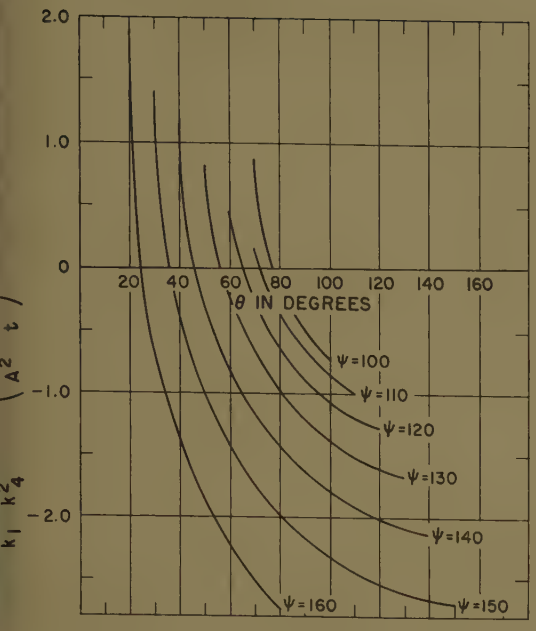


Fig. 6 (left). Variation of log $(k_2k^7)/(k_1k_4^2)$ with θ and ψ

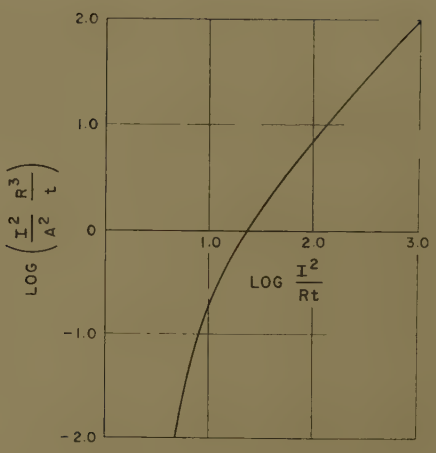


Fig. 7 (right). Transition currents for weightless drops

t should be emphasized that the sma pressures calculated are stagnation ssures, assuming no flow of plasma es. In the electric arc, the pumping ion produced by the axial pressure dient causes rapid circulation of plasma es which are introduced at the top the arc, are expanded approximately fold in volume due to arc heating, and expelled at high velocity at the base the arc. Calculation of lines of flow d pressures under dynamic flow condi- ns has not been attempted.

The axial pressure gradient

$$\left(\frac{\partial P}{\partial r}\right)_{\psi=0} = -\frac{4I^2}{\pi r^3(1-\cos\theta)^2}\ln\frac{1}{\cos\theta/2} \qquad (6)$$

l accelerate drops in transfer. This celeration has been observed in previous perimental work.[5]

Plasma pressures also exert a force on electrode tip in a direction to oppose transfer of molten metal. If it is umed that the contour of the electrode d drop is axially symmetrical and that rent flows into the plasma normal to surface of the drop, it can be proved t the opposing force is equal to $I^2/2$ nes. This force is independent of the tribution of current, provided that it axially symmetrical.

The flow of current in the necked- wn section of molten electrode metal ve the drop also induces pressure, ich exerts a downward force on the p equal to $I^2/2$ dynes. It can be ved that this force is independent of diameter of the necked-down section d the distribution of current within neck.

The external forces on the drop caused plasma pressure below the drop and hydrostatic pressure in the molten conductor above the drop are in balance. If only part of the arc current flows through the drop, the current flowing into the electrode above the drop will exert vertical forces on the drop, which also are in balance if it is assumed that this current enters the electrode normal to its surface.

To obtain the arc-induced forces, attention can be confined to the internal forces originating within the drop from the flow of arc current. To facilitate computation, it will be assumed that the drop is spherical in shape and is joined to the electrode at a profile angle ψ measured from the vertical axis. Current enters the drop through a profile angle θ also measured from the vertical axis as shown in Fig. 5.

It will be assumed that the current density at the electrode I/A is a constant for each set of welding conditions (polarity, electrode chemistry, shielding gas composition) but is independent of current.

In accordance with this assumption, the active area of the electrode must increase in direct proportion to welding current.

The net force acting on the drop as a result of the flow of welding current is (Appendix I):

$$F = I^2\left\{\ln\frac{\sin\theta}{\sin\psi} - \left(\frac{1}{4}+\frac{1}{1-\cos\theta}\right) + \frac{2}{(1-\cos\theta)^2}\ln\frac{2}{1+\cos\theta}\right\} \qquad (7)$$

where

F = force, dynes

For convenience, let k_1 equal the quantity in brackets. This parameter is a function of both θ and ψ and can be either positive or negative. Therefore,

$$F = k_1 I^2 \qquad (8)$$

When k_1 is negative (θ less than $180-\psi$), the force is negative, indicating repulsion of drops against the direction of transfer. When k_1 is positive (θ greater than 180-ψ), arc forces assist transfer.

The volume of the drop is shown in the following.

$$V = \pi R^3\left\{\frac{(1-\cos\psi)-\frac{1}{3}(1-\cos^3\psi)}{\sin^3\psi}\right\} \qquad (9)$$

where

R = the radius of the electrode, cm

Let

$$k_2 = \pi\left\{\frac{(1-\cos\psi)-\frac{1}{3}(1-\cos^3\psi)}{\sin^3\psi}\right\} \qquad (10)$$

The parameter k_2 is a function only of ψ. Therefore,

$$V = k_2R^3 \qquad (11)$$

The arc force F acting on a drop of volume V produces an apparent density ρ_e, where

$$\rho_e = \frac{F}{V} = \frac{k_1}{k_2}\frac{I^2}{R^3} \qquad (12)$$

The quantity ρ_e can be either positive or negative, depending upon the sign of k_1. Assuming weightless drops, the total density is equal to that caused by arc forces:

$$\rho_T = \rho_e$$

or

$$k_3\frac{t}{R^2} = \frac{k_1}{k_2}\frac{I^2}{R_3}$$

Note that expanding

$$\ln\frac{2}{1+\cos\theta}$$

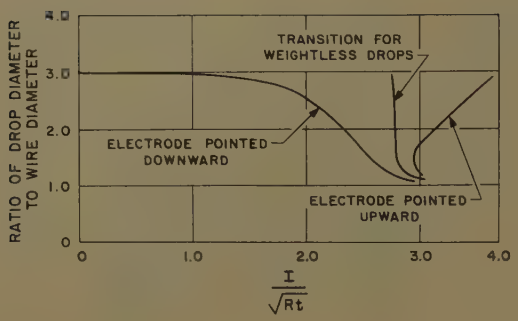


Fig 8 (left). Effect of gravity on drop size in transition region

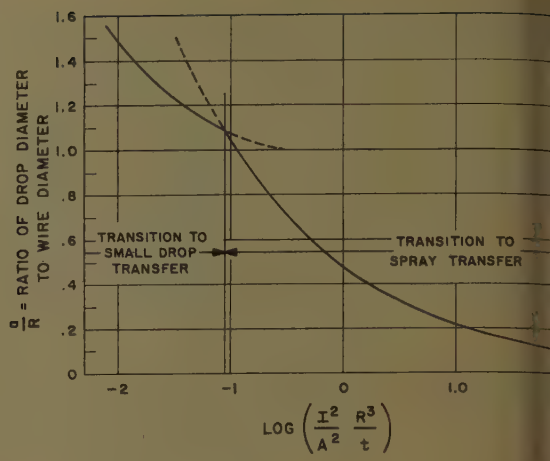


Fig. 9 (right). Drop size in posttransition region

and simplifying, k_1 is also equal to

$$\left\{\ln\frac{\sin\theta}{\sin\psi}+\frac{1}{2}\left(\frac{1}{3}\sin^2\frac{\theta}{2}+\frac{1}{4}\sin^4\frac{\theta}{2}+\frac{1}{5}\sin^6\frac{\theta}{2}\ldots\right)\right\}$$

and

$$\frac{I^2}{Rt}=\frac{k_2k_3}{k_1} \tag{13}$$

The active area of the electrode A on the drop is expressed as

$$A=2\pi R^2\left(\frac{1-\cos\theta}{\sin^2\psi}\right) \tag{14}$$

Let

$$k_4=f(\theta,\psi)=2\pi\left(\frac{1-\cos\theta}{\sin^2\psi}\right) \tag{15}$$

or

$$A=k_4R^2 \tag{16}$$

By combining equations 13 and 16

$$\left(\frac{I^2R^3}{A^2t}\right)=\frac{k_2k_3}{k_1k^2_4} \tag{17}$$

The quantity

$$\left(\frac{I^2}{A^2}\frac{R^3}{t}\right)$$

is a dimensionless group which will be designated the "transfer number" N. I/A is the electrode current density in abamperes/cm^2, and t, the surface tension, is in dynes/cm.

The transfer number N remains constant for any given set of welding conditions as current is varied, provided that the surface tension and electrode radius are not greatly affected by temperature changes resulting from this current variation. It was assumed that current density I/A is constant.

The quantity

$$\frac{k_2k_3}{k_1k^2_4}$$

is a parameter dependent only on θ and ψ. Fig. 6 is a plot of the log of this quantity as ordinate with θ as abscissa. Curves for various values of ψ are shown.

For weightless drops, the quantity

$$\frac{k_2k_3}{k_1k^2_4}$$

is equal to the transfer number N. The intersections of the characteristic curves with log N determine the drop-size–current relationship from the following:

$$\frac{a}{R}=\frac{1}{\sin\psi} \tag{18}$$

$$\frac{I}{\sqrt{Rt}}=N^{1/2}k_4 \tag{19}$$

where a is the drop radius, and k_4 is determined from the values of θ and ψ at the intersections.

It can be shown that over a range of values of a/R, the value $I/\sqrt{Rt}$ remains constant, indicating a sudden change in drop size as current reaches a critical value. This value will be called t transition current.

A plot of the log of the transition c rent dimensionless group I^2/Rt as function of the log of the transfer numb N is shown in Fig. 7.

Transition currents for weightless dr can be determined from this curve af the value of the transfer number has be calculated.

Transfer Under the Combined Acti of Surface Tension, Gravitational, and Arc Forces

Pretransition Region

When all forces are considered t total density is equal to the sum of t apparent density due to arc forces a the real density.

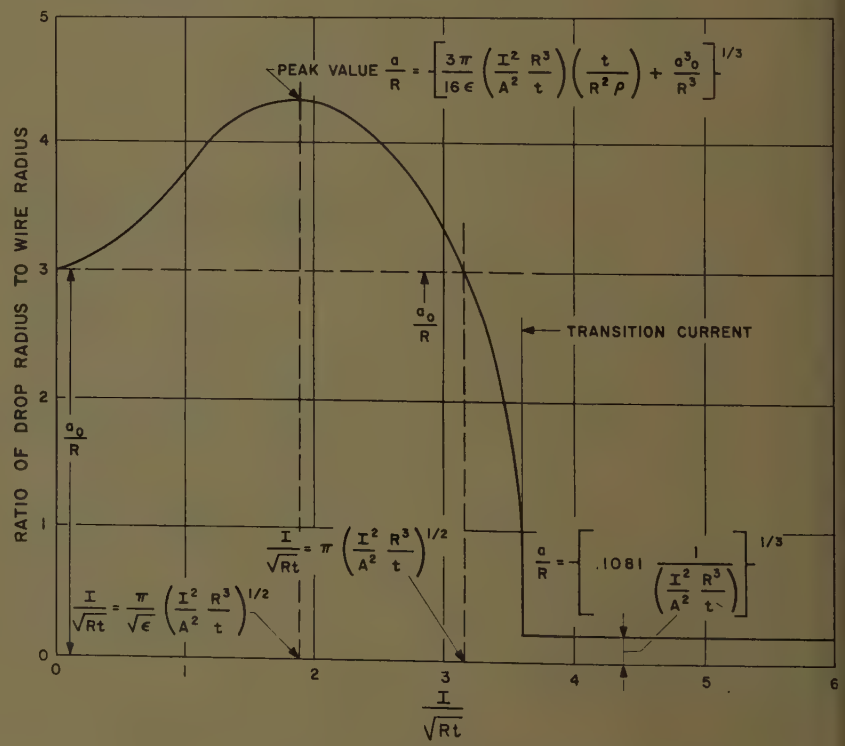


Fig. 10. Theoretical transfer curve

$\rho_e+\rho$ (20)

$=\frac{k_1 I^2}{k_2 R^3}+\rho$

ranging terms:

$\frac{k_2 k_3}{k_1}\left(1-\frac{R^2\rho}{k_3 t}\right)$ (21)

$\frac{R^3}{t}\right)=\frac{k_2 k_3}{k_1 k^2_4}\left(1-\frac{R^2\rho}{k_3 t}\right)$ (22)

$\frac{k_2 k_3}{k_1 k^2_4}\left(1-\frac{M}{k_3}\right)$ (23)

se equations apply to arcs where vity assists transfer (electrode pointed ically downward). If gravity is in osition to transfer (electrode pointed ically upward), the following equas are used:

$=\frac{k_2 k_3}{k_1}\left(1+\frac{M}{k_3}\right)$ (24)

$=\frac{k_2 k_3}{k_1 k^2_4}\left(1+\frac{M}{k_3}\right)$ (25)

r convenience, these equations can be pressed as follows:

$=\frac{N}{\left(1\mp\frac{M}{k_3}\right)}$ (26)

$\frac{k_2 k_3}{k_1 k^2_4}=\log N-\log\left(1\mp\frac{M}{k_3}\right)$ (27)

e drop-size versus current curves in pretransition region can be calculated m values of θ and ψ at points on the ves of Fig. 6. These points are deterned by plotting values of

$\frac{k_2 k_3}{k_1 k^2_4}$

culated from equation 27 on corrending ψ = constant curves.

Transition curves for weightless drops l for gravity aiding and opposing nsfer are plotted in Fig. 8. In this e, the transfer number N was assumed be 0.1, and the drop-size index M was culated for a 1/16-inch-diameter steel ctrode with ρ=7,450 dynes/cm³ and 1,200 dynes/cm.

The transition for weightless drops is y abrupt. Gravity aiding transfer duces a more gradual reduction in p size with current. When gravity oses transfer, the transition has a erse bend, indicating that a greater

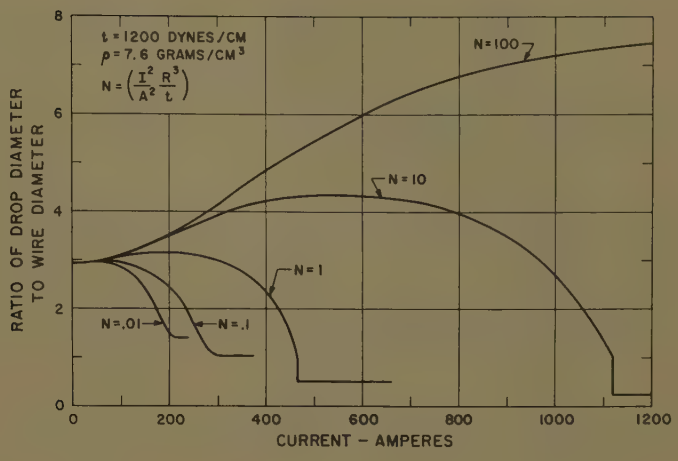


Fig. 11. Effect of transfer number on drop-size–current relationships for 1/16-inch-diameter steel electrode

current is needed to transfer large drops than is required for small drops. This is only of theoretical significance, since points above the vertical portion could not be obtained experimentally. Drops grow in size during the melting period. If current is above the critical value, small drops would transfer before reaching a diameter corresponding to points on the reverse bend.

Repulsion Region

The curves of Fig. 6 are drawn for values of ψ up to a maximum of 160 degrees corresponding to a drop diameter of approximately three wire diameters. At currents far below the transition there is a region where arc forces oppose transfer (ρ_e is negative). In this region, drops may exceed three wire diameters, necessitating a revision in calculating procedure. A close approximation to a/R can be obtained by the following equation (see Appendix II):

$$\frac{a}{R}=\left\{\frac{3}{16\pi}\left(\frac{I^2}{Rt}\right)\left(\frac{1}{M}\right)\ln\frac{\pi^2 N}{\frac{I^2}{Rt}}+\frac{a^3_0}{R^3}\right\}^{1/3} \quad (28)$$

where

$\frac{a_0}{R}$ = ratio of drop radius to wire radius in the absence of arc forces. This value can be read from Fig. 3 corresponding to $k_3=M$

The maximum value of a/R can be obtained by differentiating equation 28 with respect to $I/\sqrt{Rt}$ and setting the resultant expression equal to zero, producing the following:

$$\frac{I}{\sqrt{Rt}}=1.906N^{1/2} \quad (29)$$

and

$$\frac{a}{R}=\left\{0.2167\,\frac{N}{M}+\frac{a^3_0}{R^3}\right\}^{1/3} \quad (30)$$

Arc forces reduce to zero when the first term in brackets in equation 28 is equal to zero.

In this case,

$$\frac{I}{\sqrt{Rt}}=\pi N^{1/2} \quad (31)$$

and

$$\frac{a}{R}=\frac{a_0}{R} \quad (32)$$

Posttransition Region

The minimum value of ψ in the curves of Fig. 6 is 100 degrees, corresponding to a drop diameter slightly greater than wire diameter. With values of N above 0.1, the drop diameter in the spray region will be reduced to a value below wire diameter. In this case, the assumption of spherical drops is no longer valid. A closer approximation can be obtained by considering the drops to be hemispherical in shape at the bottom and cylindrical in shape above the hemisphere. The height of the cylinder will be assumed equal to the radius of the hemispherical end.

For values of k_3 greater than 0.85, the diameter of the drop and the diameter of the neck will be equal, in conformance with the assumed drop shape.

It can be proved (see Appendix III) that the ratio of drop radius to wire radius is inversely proportional to the cube root of the transfer number based on the mathematical model assumed.

$$\frac{a}{R}=0.4764\,N^{-1/3} \quad (33)$$

In order to validate this result, the current must be at least large enough to cover the entire surface of the drop. Or,

$$\frac{I}{\sqrt{Rt}}\geqq 2.852\,N^{-1/6} \quad (34)$$

Equation 33 applies to the spray region where drop diameter is less than wire diameter. In the small-drop region, values of N where $\theta=\psi$ on each curve of Fig. 6 determine the minimum size of

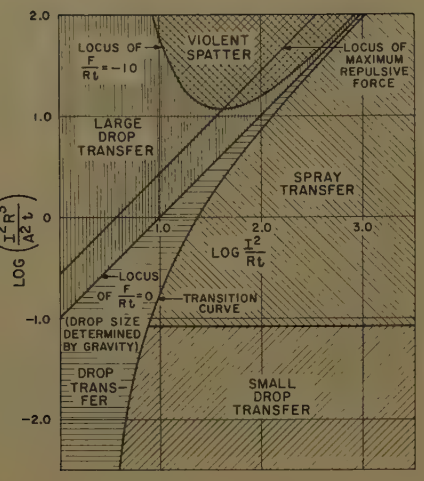


Fig. 12. Transfer chart

drop after the transition. Any further increase in current merely causes the active electrode area to extend upward on the electrode wire above the drop, which (according to the mathematical model assumed) has no effect on drop size.

The two approximations to drop size in both the spray and small-drop regions are plotted in Fig. 9. The intersection of these two functions establishes the value of N, separating the small-drop and spray areas of the posttransition region. The general shape of the drop-size–current function in dimensionless form is shown in Fig. 10. Formulas to locate key points by means of the transfer number N, and drop size index M, are shown.

Using these formulas, drop-size–current functions for various values of the transfer number N are plotted in Fig. 11. In calculating these theoretical curves, a 1/16-inch-diameter steel electrode was assumed (density $\rho=7{,}450$ dynes/cm^3 and surface tension $t=1{,}200$ dynes/cm). Note that the transition to small-drop transfer with $N=0.01$ and 0.1 is smooth, while an abrupt reduction in drop size occurs at the transition from drop to spray transfer.

A transfer chart (see Fig. 12) was constructed to assist in visualizing the effects caused by changes in the transfer number N. The locus of points where arc forces reach maximum repulsion (maximum drop size); the locus of points where arc forces change in direction from repulsion to attraction (drop size determined by gravity); and the locus of transition currents are plotted.

Lower values of N reduce transition current, while higher values produce strong repulsion in the useful welding current range, causing wild spatter.

Transition currents for values of N above 1 (log $N=0$) are usually above the useful welding current range. A transfer number from 0.01 to 1 will produce satisfactory welding (log $N=-2$ to 0).

Discussion of Theory

The theory discloses that in arcs where the electrode current density is excessive, transfer will deteriorate. However, if current is raised manyfold, there will be a transition to spray transfer with very fine drop size. Unfortunately, the methods for measuring electrode current density are very crude. Greater accuracy in determination of current densities at the surface of an electrode may be obtained by measuring transition currents and reading the corresponding value of N from which the term I^2/A^2 can be deduced.

It is postulated as a result of this theory that wild spatter produced by many welding conditions using straight polarity can be traced to the high-current densities at the cathode of the cold-cathode arc. This is in agreement with values reported in the literature where cathode current densities are at least one order of magnitude higher in cold-cathode than in thermionic electrodes. Transfer in reverse polarity arcs is usually satisfactory because of the lower current densities at the anode. The effects of shielding gas composition on transfer may also be traced to the variations they produce on the current density at both anodes and cathodes.

An increase in wire size will increase the transition current in the welding arc. The effect is produced not only by the larger transfer number N, but also by the change in scale of the current axis in the chart of Fig. 12. If wire radius is doubled, maintaining all other parameters constant, the new operating point on the chart will be displaced 3 log 2 in an upward direction and log 2 to the left.

An increase in surface tension will also increase the transition current. When surface tension is increased by a factor of 2, the new operating point will be displaced downward and to the left by log 2 in each direction.

Methods for Improving Transfer

One of the most effective methods for improving transfer in straight-polarity and a-c arcs is the use of emissive-agent-coated electrode wire.[6] In this case, formation of an atomic film thermionic cathode reduces the current density an order of magnitude below that of the bare-metal cold-cathode arc. The consequent reduction in transfer number N lowers the transition current to a value within the welding range. Other methods of inducing thermionic arc action similarly effective.

Another method, called "dip transfe has been effective not only in eliminat spatter in carbon-dioxide-shielded reve polarity welding of steel, but also depositing lower temperature weld me Vertical and overhead welding is fac tated by this effect.

In dip transfer welding, a const potential or rising characteristic po supply with a controlled amount of l circuit inductance is adjusted to an o put voltage slightly below arc volta With this setting, a sustained open cannot be maintained. Periodic sho circuiting of the arc stores energy in inductance needed to raise output v age to arc voltage level during the arc period. Inductance is adjusted to n enough wire to form a small drop. decay of current during the arcing per reduces repulsive arc forces and burn rate so that short-circuiting can ta place. Surface tension then transfers drop into the weld pool. This meth permits spatterfree welding even when transfer number is too large for sa factory transfer.

Conclusions

Transfer in welding arcs may be ch acterized by two dimensionless para eters: M (drop size index) and N (tra fer number). For satisfactory weldi the transfer number should lie within range 0.01 to 1. By means of th parameters, the transition current, as w as the drop size in the repulsion, pretra tion, and posttransition regions, may determined.

It should be emphasized that the the in its present form will apply only to a in which the current density at the act surface of the electrode is uniform a remains constant with changes in curre This condition is not satisfied in jet re tion arcs and in partially stabili straight-polarity arcs where both cold a thermionic cathode action are occurr simultaneously. It was also assum that the surface tension, density, electrode diameter remain constant w changes in temperature. Variations produce correspor ding variations in values of M and N over the weld current range.

All phases of this theory have not b subjected to experimental verificati However, the variation of transition c rent with wire size shows good agreem with measured values. Drop size in spray region is somewhat greater t measured values. This may be att

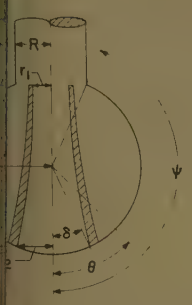


Fig. 13. Incremental section of integration of arc forces

to the fact that the current may flow into the plasma normal to the ten electrode surface as assumed. ps may then form in the decreasing sure gradient of the arc plasma, ing e greater than the calculated e.

n general, the theory leads one to the clusion that deterioration in transfer er certain conditions of welding can be ibuted to excessive current density at electrode surface. Final vindica- of this point of view depends upon results of further research into anode cathode phenomena and the factors ch control current density. An in- al part of this effort is the development nore accurate methods for the measure- nt of electrode current densities.

Appendix I. Pretransition Region

spherical drop is attached to a wire at angle ψ from the axis (Fig. 13). Assume t the drop carries current uniformly dis- uted in the wire at the point of attach- nt. Current flows with axial symmetry ough the drop and into an arc plasma ough an angle of rotation θ.

Consider an element of the drop with al symmetry constructed so that the er and outer profiles lie along current amlines separated by an incremental nge in radius.

The current flow within the element is stant due to the method of construction. the value of this current in abamperes al di. Let the current enclosed by the nent in abamperes equal i. If r is the ial distance from the axis to any point he element in centimeters,

$$\frac{\mu 2_i}{r}$$

ere

flux density, gauss

The downward component of the "motor ion" force produced in a section of the ele- nt of length dL is as follows:

$=B\ di\ dL$ in ϕ

ere ϕ is the angle formed by the tan- t to the surface of the element and the .

$$dF=\frac{2\mu_i}{r}\,di\,dL\sin\phi$$

but

$$\sin\phi=\frac{dr}{dL}$$

or

$$dF=\frac{2\mu_i}{r}\,di\,dr$$

$$F=\int_{r_1}^{r_2}\frac{2\mu_i}{r}\,di\,dr$$

or

$$F=2\mu_i\,di\ln\frac{r_2}{r_1} \qquad (35)$$

where r_1 and r_2 are the radii at the upper and lower extremities of the element.

This result is significant since it proves that the downward component of the force on the element is independent of the particular path taken by the current. The radii at the points of current entry and exit are the values of interest.

Integrating over-all elements of the drop,

$$\text{total } F=2\mu\int_0^I\left(i\ln\frac{r_2}{r_1}\right)di \qquad (36)$$

where

I = total current in the drop, abamperes

but

$$\frac{\pi r_1^2}{\pi R^2}=\frac{i}{I}=\frac{2\pi a^2(1-\cos\delta)}{2\pi a^2(1-\cos\theta)}$$

$$r_1^2=R^2\frac{i}{I}$$

$$r_2^2=a^2\left\{2\frac{i}{I}(1-\cos\theta)-\left(\frac{i}{I}\right)^2(1-\cos\theta)^2\right\}$$

or

$$\left(\frac{r_2}{r_1}\right)^2=\frac{a^2(1-\cos\theta)^2}{R^2}\left\{\frac{2}{(1-\cos\theta)}-\frac{i}{I}\right\}$$

Substituting in equation 36

$$F=2\mu\int_0^I i\left\{\ln\frac{a(1-\cos\theta)}{R}+\frac{1}{2}\ln\left(\frac{2}{1-\cos\theta}-\frac{i}{I}\right)\right\}di$$

Since $\mu=1$ for metals above the Curie points,

$$F=I^2\left\{\ln\frac{\sin\theta}{\sin\psi}-\left(\frac{1}{4}+\frac{1}{1-\cos\theta}\right)+\frac{2}{(1-\cos\theta)^2}\ln\frac{2}{1+\cos\theta}\right\} \qquad (7)$$

Appendix II. Repulsion Region

In the repulsion region

$$k_1\cong\ln\frac{a\theta}{R}$$

and

$$I^2\cong N\frac{t}{R^3}\pi^2a^4\theta^4$$

or

$$a^4\theta^4=\frac{R^3}{\pi^2Nt}I^2$$

$$F=k_1I^2$$

$$F\simeq\frac{I^2}{4}\ln\frac{I^2}{\pi^2NRt}$$

Setting repulsive force equal to increase in drop weight,

$$-F=\frac{I^2}{4}\ln\frac{\pi^2NRt}{I^2}=4/3\ \pi\rho(a^3-a_0^3)$$

where

a_0 = radius of drops determined by gravity and surface tension

Solving for a^3 and dividing by R^3

$$\frac{a^3}{R^3}=\frac{3}{16\pi}\left(\frac{I^2}{Rt}\right)\left(\frac{1}{M}\right)\ln\frac{\pi^2N}{I^2/Rt}+\frac{a_0^3}{R^3}$$

or

$$\frac{a}{R}=\left\{\frac{3}{16\pi M}\left(\frac{I^2}{Rt}\right)\ln\frac{\pi^2N}{I^2/Rt}+\frac{a_0^3}{R^3}\right\}^{1/3} \qquad (28)$$

Appendix III. Spray Area

Assume that maximum-size drop can be represented in shape by a hemisphere of radius a joined to a cylinder of radius a and height a.

Internal force, spherical section

$$=\frac{I^2}{4}(\ln\sqrt{2}+0.1363)$$

Internal force, cylindrical section

$$=\frac{I^2}{2}\left(\frac{3}{8}-\frac{1}{4}\ln 2\right)$$

Total force $F=0.2216\ I^2$

$$V=\frac{2}{3}\pi a^3+\pi a^3$$

$$\rho_e=\frac{F}{V}=\frac{0.1330}{\pi}\frac{I^2}{a^3}$$

but

$$I^2=16\pi^2a^4N\frac{t}{R^3}$$

$$\rho_e=2.127\,\pi\frac{Nat}{R^3}$$

$$\rho_T=\rho_e+\rho$$

$$0.7225\frac{t}{a^2}=2.127\pi\frac{Nat}{R^3}+\rho$$

$$\frac{a^3}{R^3}=\frac{0.7225}{2.127\pi}\frac{1}{N}-\left(\frac{R^2\rho}{t}\right)\frac{1}{2.127\pi N}\frac{a^2}{R^2}$$

For wire sizes normally used in gas-shielded welding, M is less than 0.1.

$$\frac{a}{R}\cong 0.4764N^{-1/3} \qquad (33)$$

References

1. Forces Acting During the Transfer of Material Through the Welding Arc, J. Sack, Jr. *The Welding Industry*, London, England, July 1939.

2. Forces of Electric Origin in the Iron Arc, F. Creedy, R. O. Lerch, P. W. Seal, E. P. Sordon. *AIEE Transactions*, vol. 51, June 1932, pp. 556–66.

3. Overhead Welding, J. Sack, Jr. *Phillips Technical Review*, Eindhoven, Netherlands, vol. 3, no. 1, Jan. 1939.

4. Differential Gleichungen, E. Kamke. Akademische Verlagsgesellschaft m.b.h., Leipzig, Germany, pp. 170–71.

5. Metal Transfer Characteristics in Gas Shielded Arc Welding, H. C. Ludwig. *Welding Research Supplement*, London, England, Jan. 1957.

6. The Atomic Film Cathode in Inert G Shielded Metal Arcs, A. Muller, W. J. Gree "Conference on Electric Arc and Resistance We ing—IV", *Special Publication S-64*, July 1954, 61–74.

7. Method of Graphic Computation of Su face and Interfacial Tension by the Form o Stationary Drop, S. I. Popel, et al. Tru Uralvskogo Politekhnicheskogo Instituta, Sve lovsk, USSR, vol. 49, 76-B1, 1954.

Discussion

A. A. Smith, A. A. Wells, J. C. Amson (British Welding Research Association, Abington, England): The invitation by Mr. Greene to submit a discussion of his paper has been warmly received, particularly since the subject of the paper is currently receiving attention at the British Welding Research Association, and the nature of the difficulties in preparing such an analysis is appreciated. In making use of dimensional analysis, the author has achieved a commendable separation of the many pertinent variables into dimensionless groups and has allowed attention to be focussed on those variables which apparently dominate the mechanism of metal transfer in arcs. May we, in passing, point out that the author uses ρ and density when, in fact, ρg and specific weight are often intended.

Investigation at the British Welding Research Association has disclosed the necessity of deciding between two alternative hypotheses of current emission into the arc plasma from the droplet forming on the wire electrode. Categorically these are (1) that the current emerges into the plasma at a constant surface density j, the activated electrode area varying in direct proportion to the current; and (2) that the current emerges uniformly over all the fused surface of the droplet during its formation. High-speed photographic observation cannot be relied upon to discriminate between these two hypotheses, since inevitably the observation of the plasma's disposition about the electrode is an observation of radiation in the visual and ultraviolet frequencies and not of current conduction. An additional complication is that the current may not emerge uniformly at all; the greater mobility of electrons as compared with ions may allow an axial concentration of electron current in response to the Lorentz forces which determine conduction paths in the plasma. However, we have been greatly interested to see that the author chooses the first hypothesis of j = constant, and to observe the consequences in the form taken by the analysis and the resulting prediction of behavior in the system.

It is clear that the second hypothesis is not tenable for low currents emerging from large drops, but it is in just this range of conditions that the system is least sensitive to current-induced forces, transfer being principally dependent on gravitation. Using the second hypothesis for dimensionless currents ranging between about one-half and full transition value, we predict a behavior closely similar to the author's. This prevents a rejection of one or other of the hypotheses. Similarly, a prediction of spatter occurs using either hypothesis, providing only that the absolute value of the current is high enough and the droplet diameter large enough to give rise to an effective repulsive force. In parentheses, we were most interested in Mr. Greene's conclusion that spatter is producible by hydromagnetic forces in the droplets alone, and observe that, in addition to a repulsive force occurring whenever, in his theory, the diameter of the current emission area on the droplet is less than the diameter of the wire, it also occurs whenever, in both the author's and our theory, the current leaves the entire fused surface of a spherical droplet whose diameter is greater than about twice that of the wire. The conclusions given in the paper have also drawn attention to this possible feature of metal transfer; namely, that should for some reason certain conditions and materials promote the formation of nonconducting films over the droplet's fused surface, then the repulsive force arising from high-density current emission from small punctures in this film could be as large as one likes, providing only that the punctures are small enough. This is an aspect of transfer which deserves more investigation.

To continue, in the posttransition range the two hypotheses become effectively equivalent insofar as the author permits his current to emerge uniformly over both the hemispherical cap and the cylindrical body of the posttransition droplet. The analytical equivalence breaks down, however, when he fixes the length of the molten cylindrical body of the droplet equal to the radius of the droplet and allows current emission, in excess of that which can be borne by the fused surface at the agreed density j, to take place ineffectively from the solid wire surface above the fused droplet.

We continue to postulate that current emerges but only from the whole fused surface and that the length of the fused surface is determined otherwise and not necessarily made constant and equal to wire radius. We would suggest from the similarity of the results of these two approaches that at present both these hypotheses appear equally acceptable in enabling predictions of behavior to be made for correlation with experimental results. However, the simpler analysis following the second hypothesis throws some doubt on the advisability of introducing at this stage Mr. Greene's somewhat complicated analysis following from the introduction of his transfer number N. It is nonetheless agreed, and he is to be congratulated on his successful handling of the difficulties of his analysis, that if the first hypothesis is made, then the appearance of N is inevitable.

At the time of receipt of the paper, a droplet transfer diagram, of the type shown in Fig. 10, derived from our own analy had already been compared with the resu of welding experiments with argon shieldi using copper, steel, and aluminium wires several diameters, and the agreeme reached between theory and experiment w very encouraging. While the form of t relationship deduced by us is essentially t same as the author's, we have shown th the dimensionless transition current shou be about 2, to agree with the best expe mental determinations of clean liquid me surface tension values, and not as high the value of 3.6 deduced in the paper. T implications stem from this: (1) that th is a sufficient discrepancy between t values for surface tension derived from t cleanest of laboratory techniques and th likely to be encountered in practical we ing experiments, as to make the form values of questionable validity; and (that the hydromagnetic forces assisti droplet detachment are in fact higher th those accounted for by Mr. Greene.

Attention has been given to this lat point which is foreshadowed in the secti in the paper on weightless transfer, and c conclusions suggest that the entrainme of shielding gas into the arc plasma by hydromagnetic pumping action may be sponsible. It is sufficient to suggest th the hydromagnetic action in the gas fi surrounding the droplet may produce vel ity rather than static head at this positi so that the pinch force $I^2/2$ is not who reacted. In this event, the droplet deta ment force could be increased by the or of this amount. In fact, an elementa treatment of the hydromagnetic pumpi action has been produced whereby calc lated arc forces correlate closely with pu lished measured values. These values a greater than those calculated in the pap Early publication of our own work is an cipated.

In conclusion, we are indebted to M Greene both for the opportunity to disc his paper, and in a general way for t stimulus his paper has given to our own vestigations.

Thomas B. Correy (Hanford Atomic Pro ucts Operation, General Electric Compa Richland, Wash.): The paper is a very f study of the phenomenon of metal trans in the metal inert-gas-shielded arc and t arc phenomenon. Interest in the paper widespread, as it explains very clearly t mechanisms and limits of the various ty of metal transfer and parameters affect transfer.

W. J. Greene: I wish to express my app ciation to Dr. Smith and his associates

eir comments and suggestions. It is ıe that the symbol ρ has been used roughout the paper as a body force in nes/cm^3 rather than as a density. I am lebted to the discussers for pointing out is inaccuracy.

Dr. Smith proposes two alternative hytheses of current emission into the arc asma from the droplet forming on the re electrode. It is interesting to note at both of these hypotheses lead to the ne general conclusions regarding metallic ansfer. However, it is not clear how the ond hypothesis leads to a region of repule arc force when the drop diameter exds approximately two wire diameters. it is assumed that the current flows from e drop into the plasma normal to the sure of the drop, the arc forces under the ond hypothesis can be determined from uation 7 by setting angle θ equal to gle ψ. If this is done, the first term in ckets ($\ln \sin \theta / \sin \psi$) is equal to zero. nder these conditions it can be proved that the factor k_1 is positive for all values of θ over the entire range of possible values from zero to π. If k_1 is positive, the arc forces must assist transfer. It is clearly evident in motion pictures of the arc that under certain welding conditions there exists a strong repulsive force opposing transfer. As Dr. Smith suggests, the formation of nonconducting films over the droplets' fused surface may produce current conduction through small punctures in this film, and in that case repulsive forces could be unlimited. However, this particular feature of transfer is not encompassed in either of the hypotheses mentioned. It would also appear that the second hypothesis leads one to the conclusion that transfer characteristics on straight and reverse polarity should be identical. Furthermore, the differences in transfer which occur with various shielding gases can be explained only on the basis of the corresponding changes in surface tension that they produce. This effect is minor in relation to the gross changes which can be detected by experimental means. The first hypothesis can explain both of these effects by the difference between anode current density and cathode current density, and the variations in these current densities which are produced by shielding gas composition.

I note with interest the conclusion that entrainment of shielding gas into the arc plasma by its hydromagnetic pumping action may aid transfer by reducing the static head at the base of the drop. A mathematical treatment of this effect would be an important contribution to the theory of transfer.

I wish to express my appreciation to Dr. Smith and his associates for introducing several stimulating ideas which deserve more thorough consideration than has been devoted to them at this time. I look forward to the publication of their work on metallic transfer.

I also wish to thank Mr. Correy for his comments.

A New Welder Busway Distribution System

LAWRENCE E. FISHER
MEMBER AIEE

ROBERT W. DAILEY
ASSOCIATE MEMBER AIEE

THE OHIO Stamping Plant of Chrysler Corporation at Twinsburg, Ohio, vers an area of 1,500,000 sq ft (square et), approximately 34 acres. Power is pplied by the Ohio Edison Company to pply the great number of motor loads d so forth for the heavy presses and her manufacturing equipment in the ant. The estimated short-circuit duty 2,500,000 kva, based on long-range ture plans. The estimated initial shortrcuit duty is 600,000 kva 3 phase symetrical.

Power at 138 kv is supplied to three ansformers rated 10,000/12,500 kva, phase, 60 cycles, 138 kv–13.8 kv, conected delta-Y with secondary grounded. l of the resistance welding discussed in is report is supplied from one of the ,000/12,500-kva transformers provided with a voltage regulator. The other transformers supply the press motors and other loads in the stamping plant.

The 13.8-kv power is supplied to the metalclad switchgear through a 1,200-amp (ampere) breaker. Feeder breakers are also rated 1,200 amp.

From the feeder breakers in the metalclad switchgear, power is supplied through runs of 3-conductor 350-MCM (thousand circular mils) cable to 14 2,000-kva transformers in seven double-ended unit substations as shown in Fig. 1. The transformers are rated 2,000 kva, 3 phase, 60 cycles, 13.8 kv–480 volts, ungrounded secondary, connected delta-Y. Also there are 10 1,000-kva transformers, supplying other than welding loads.

The entire plant area is covered with networks of approximately 13 miles of different types of busways. Light-duty 50-amp plug-in busways serve the fluorescent lighting load. Plug-in busways in ratings of 400 and 600 amp, having aluminum bus bars, braced for a 50,000-amp short-circuit current serve the motor loads and general factory requirements. New 1,600-amp, type *LVDP*, low-voltage-drop plug-in busways supply the heavy welding loads.

Three double-ended unit substations, each with 2 2,000-kva transformers and with 3,000-amp main and tie circuit breakers, supply the heavy welding departments. Each transformer feeds two runs of the 1,600-amp type *LVDP* busway each protected by 1,600-amp feeder circuit breakers. The runs are approximately 400 feet long, with the outer end of each run connected to the end of an-

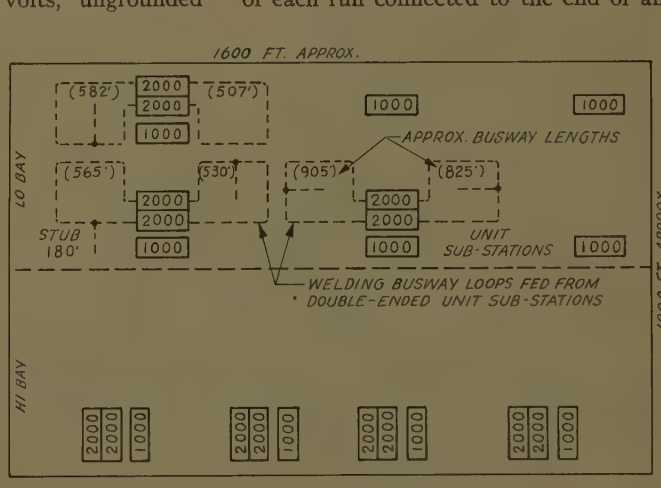


Fig. 1. Physical location of unit substations and welder busways at Ohio Stamping Plant

per 60-140, recommended by the AIEE Electric elding Committee and approved by the AIEE chnical Operations Department for presentation the AIEE Winter General Meeting, New York, Y., January 31–February 5, 1960. Manuscript mitted September 24, 1957; made available printing November 27, 1959.

WRENCE E. FISHER is with the General Electric mpany, Plainville, Conn. ROBERT W. DAILEY with the Chrysler Corporation, Twinsburg, Ohio.

e authors wish to acknowledge the assistance of n Hamilton, R. H. Kaufmann, W. E. Long, l Heinrich Hansen of the General Electric mpany, and of V. C. Aderente of Albert Kahn sociated Engineers.

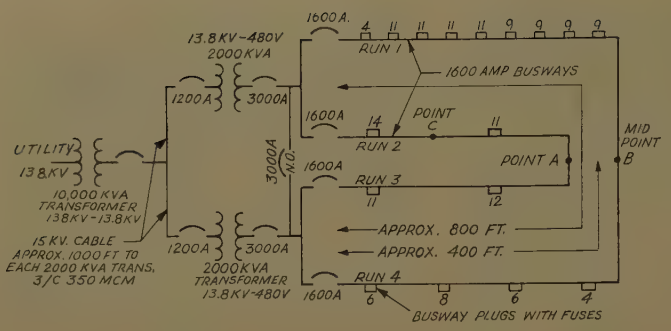


Fig. 2. One-line diagram of part of welder distribution system showing busways fed from one of the unit substations

other run to form an 800-foot loop between 2 of the 2,000-kva transformers. Actually, the busway loops vary from 507 to 905 feet in the total length. There are six welder busway loops. Thus the transformers are effectively tied in parallel on the 480-volt side by two parallel runs of 1,600-amp busway.

This paper deals primarily with these three double-ended unit substations and the 12 runs (or 6 loops) of 1,600-amp busways feeding the heavy welding departments. These are shown in Figs. 1 and 2.

Method of Selecting System Elements to Feed the Resistance Welder Loads

A great amount of preliminary work was done in selecting the elements of the distribution system to feed the welder loads in this plant. As is customary on many jobs of this magnitude, the preliminary design was based primarily on previous experience and know-how.

After making the preliminary selection of the system elements on the basis of this previous experience, it is then customary to make careful calculations covering all phases of the problem. This is done in order to make final selection of system elements that will be adequate not only to satisfy the thermal requirements, voltage-drop requirements, and short-circuit requirements, but also to select protective devices that will co-ordinate properly, providing as much selectivity as possible and providing high-speed tripping for both the maximum possible values of short-circuit current and also for the lower values of arcing fault current.

This part of the paper is intended to show by an example how to select elements of the distribution system to feed welder loads, satisfying all these requirements.

Starting with the magnitude, location, and characteristics of the welder loads, the selection of the elements of the distribution system proceeds as follows, assuming no previous field experience to aid in the selection.

The welding load consists primarily of approximately 500 65-kva single-phase welding transformers totaling about 36,000-kva nameplate rating.

The multitransformer welding presses are similar to the type shown in Fig. 13 of reference 1, and the characteristics of the load are about as described in the same reference. The power factor is relatively high, about 70%.

In order to define the "during-weld" demand, it was necessary to consider the probability of several welder groups firing at the same time. Two different methods of firing the welders were considered, but after a preliminary study it was decided to fire a group of welding transformers simultaneously 3-phase, with one third of them connected to each phase of the 3-phase system. This eliminates the problem of dealing with single-phase currents drawn from a 3-phase system.

A preliminary study was made for the welding control to fire the welders single-phase by first firing a group of 4 welding transformers on phase AB for 15 cycles, for example, then 4 on phase BC for the next 15 cycles, then 4 on CA, etc. This method has an advantage, in that a smaller electrical distribution system can be used, but the unbalance of the 3-phase system was considered undesirable.

To provide as much flexibility as possible, a type of welding control was selected that can be used to fire the welders either single-phase or 3-phase. For this example, however, the welders will be fired 3-phase.

A close study of different methods of welding indicates that the design of the system depends more on the maximum number of welders fired simultaneously than on the average number. This is due to the fact that even with a fairly large number of welders, the number of welders which will be firing simultaneously based upon a given probability is relatively small. With this small number of overlapping welders, it is very probable that the welders that are welding simultaneously during the same 15-cycle interval may all be the ones having the maximum number of welding transformers. Thus, regardless of the method used, it would be desirable to make an effort to reduce the maximum number to be fired simultaneously, especially if there are only a few groups containing the maximum number.

For example, the following is a list of the number of transformers fired simultaneously in the different stages of welding the front and rear floor pans. Since no more than 14 are fired together except in the 2nd, 3rd, 4th, and 5th stages, it is suggested that these stages be split in two, firing half the transformers in one 15-cycle period and the other half in the following 15-cycle period. This reduces the maximum number from 22 to 14 and the average in the front floor pan, which is the heaviest loaded welding operation in the plant, from 12 transformers per firing to 8.3.

Welding Transformer Lineup: Ohio Stamping Plant

Floor Pan	Operation, Stages	No. of Transformers
Front	1	4
	2	22
	3	22
	4	18
	5	18
	1*	4
	2*	6
	3*	8
	4	6
		108
Rear	1	14
	2	11
	3	12
	4	11
		48
Total		156

* Table top.

The heavy welding departments are supplied by three double-ended unit substations, each having 2 2,000-kva transformers with a main circuit breaker for each transformer, a tie circuit breaker and feeder circuit breakers. Each unit substation is centrally located with respect to a number of the multitransformer presses; see Fig. 2.

The 65-kva welding transformer loads are in groups of a maximum of 14, 5 phase, fed from each fused switch busway plug. The maximum of 14 transformers are fired simultaneously by Ignitrons. Thus, the total number of groups in run 1 is 9, in runs 2 and 3 is 2, and in run 4 is 4.

A test indicates that the during-weld current of each 65-kva welding transformer is 250 amp. All welders will weld for 15 cycles out of each 9 seconds, thus, the duty cycle is 0.0277.

In a consultation with the plant engineers it was agreed that it is desirable not to have more than 1 bad weld per 10,000, or 1/100 of 1%. Thus the system

gn should be based on a probability bad weld per 10,000.

is decided, further, to standardize way sizes using the front and rear floor department as a basis, since it is now most heavily loaded. The standard- ion is desired because of the ever- nging conditions in the automotive stry. The number of welding trans- ers fired simultaneously at any loca- can be changed easily. With two way runs 400 feet long from each of the ransformers in each unit substation hown in Fig. 2, if the ends of the bus- s at the extreme right are connected ther, as shown, the transformers will n be connected in parallel by the two allel 800-foot loops of busways. Such arrangement has several advantages, nay be seen in Appendix I.

. With a single large load connected r the input end of either busway the transformers will divide the current the ratio of approximately 60% and %. Thus, with almost any practical ding around the loops, the trans- ners will divide the load almost ally.

. As a result of this sharing of the d the voltage drop in the transformers l be reduced greatly.

. Any single large load, or center of eral distributed loads, located near mid-point of one loop, such as at nt A or B in Fig. 2, will result in about ual currents flowing in the two halves the loop. Any load at point C, at the d-point of one run (one fourth of the p), will result in a division of current 65% and 35% in the two parallel paths, ound the loop.

r Each Part of the System, It Is Necessary to Make Two Separate Calculations

1. *Thermal Requirement.* First, a cal- ation is made to determine the ampere ing of the fuse, busway, or other ele- nt to satisfy the thermal requirement. is is sometimes called the equivalent ntinuous current. For each group of lders use the equation:

$$I_d\sqrt{\text{duty cycle}}$$

which: I is the ampere rating required, is the total during-weld current for all welders fired simultaneously in the up. Select a fuse or molded-case cir- t breaker about 25% larger than the pere rating (I) calculated in order to vent nuisance blowing or tripping. is safety factor is not necessary for the sway, transformer, or magnetic circuit aker although some engineers apply 1.25 factor without it being clear whether the factor is or is not at least partly for future growth.

To determine the total equivalent continuous current[2] to a number of welders connected to one busway, use the following equation, in which I_t is the total equivalent current, d is the duty cycle, and $I_1, I_2, \ldots$ etc., are the during-weld currents in the different welders or groups of welders.

$$I_t = \sqrt{(d-d^2)(I_1^2+I_2^2+\ldots)+d^2(I_1+I_2+\ldots)^2}$$

If the duty cycles of the welders are not the same, use the equation

$$I_t = \sqrt{I_1^2(d_1-d_1^2)+I_2^2(d_2-d_2^2)+\ldots+(I_1d_1+I_2d_2+\ldots)^2}$$

2. *Voltage Drop Requirement or Total During-Weld Requirement.* Second, a calculation is made to determine the total maximum during-weld current. This takes into consideration the probability of several welders firing simultaneously.

Taking each fuse, circuit breaker, busway or other element one at a time, and, unless conditions are identical, preferably starting at the smallest branch units, such as the fuses in the busway plugs, determine the total number of welding transformers fed through that fuse; then, using the proper probability[3] curve from Fig. 3, determine the number of simultaneous welds. Use the bottom scale. Obviously the curves in Fig. 3 would not be required if all welders are fired simultaneously by Ignitrons. In such a case, take the total number of welders. In either case, multiply this number by the during-weld current of one welding transformer to determine the maximum total during-weld current I_t passing through the fuse or other element. Use maximum values of during-weld current, whereas for the thermal requirement an average value may be used.

Then, for the fuse or circuit breaker previously selected to satisfy the thermal requirements, check its time-current curve to make sure that it will not blow, or trip, for this total during-weld current I_t in the time that the during-weld current flows. Increase the ampere rating of the fuse or circuit breaker if necessary. Note that, in the example to follow, a safety factor is applied to prevent nuisance tripping. Tripping of circuit breakers, even though it occurs only once or twice a week, is very undesirable. This safety factor is of importance primarily for thermally sensitive devices such as fuses and molded case circuit breakers which may be at normal operating temperature at the time the welding surge is applied.

For the busways, cables, or transformers previously selected to satisfy the thermal requirements, calculate the voltage drop for this total during-weld current I_t.

I_t is calculated using the probability curves, the same as for fuses or circuit breakers. It is necessary, however, to subtract one from the total number of connected welders before using the curves, because the voltage drop or variation at any welder is due only to variation in current flowing to other welders. Use the upper scale in Fig. 3. There is no variation due to the current flowing to the welder under consideration because every time it fires it draws the same current. Obtain from the curve the number of welders firing simultaneously. Then multiply by the during-weld current of

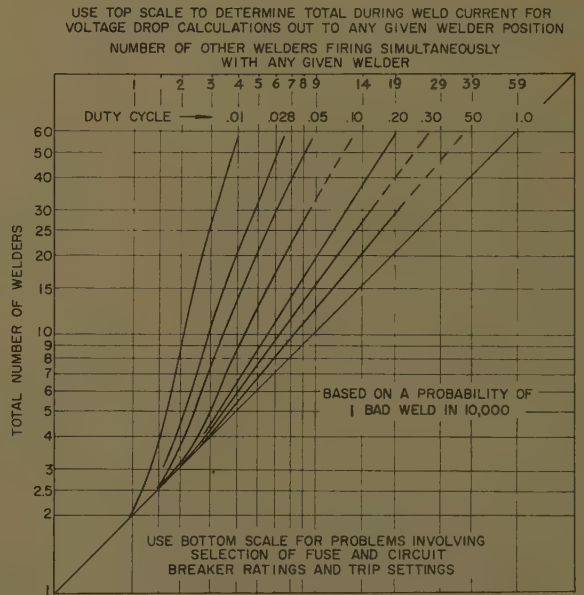


Fig. 3. Probability curves for determining number of welders firing simultaneously

Table I. Recommended Fuse Sizes

Number of Welding Transformers Connected to One Busway Plug	During-Weld Current I_d, Amperes	Equivalent Continuous Current I, Amperes	Recommended Fuse Sizes
3	433	72	100
6	866	144	200
9	1,300	216	300
12	1,732	288	400
15	2,166	360	450

one welder, or group of welders, to determine the total during-weld current for the voltage-drop calculation.

If the voltage drop is too high, a larger busway or transformer would then be selected. In general, the voltage drop should not exceed 15%, although in critical work a lower limit is required.

In the curves of Fig. 3, note that half any given number of welders results in more than half the corresponding number of simultaneous welds. Thus, if a main feeder has two equal branches, the number of simultaneous welds for the branches is more than half that for the feeder.

With this preliminary information it is now required to determine the sizes of the fuses in the busway plugs, the size of the busways and protective feeder circuit breakers, and the size of the main circuit breakers and transformers.

Solution of Problem: Thermal Requirement

First, starting at the smaller branch elements out near the load, make a selection of the busway plugs and fuse sizes, followed by the busways, feeder breakers, main breaker and, finally, the transformer.

Fuses in Busway Plugs

For each group of 3 65-kva transformers, 1 per phase, the single-phase during-weld current is 250 amp or 433 amp 3-phase = I_d. The current based on the thermal requirement $I = I_d \sqrt{\text{duty cycle}} = 433 \times 0.166 = 72$ amp. Recommended fuse sizes are shown in Table I. This assumes 433 amp as the average current.

Busway No. 1 with No Tie Busway at End

The ampere rating required for the busway is 72 amp for each group of 3 transformers. The total of 84 transformers or 9 groups would require a busway rating of 790 amp calculated by the method developed in reference 2; see Fig. 2.

Busway No. 4 with No Tie Busway at End

Seventy-two amp for each group of 3 transformers amounts to approximately 342 amp for the 24 transformers connected to Busway 4 using the same method.

Busways 1 and 4 Forming a Loop (Connected at Point B)

Although an 800-amp busway could be used for Busway 1 and a 400-amp for Busway 4, it was previously decided to standardize the busway sizes and use the double loop system. By preliminary voltage-drop calculations (conducted later) similar to those described in the section entitled Voltage Drop of Busway Loop 1-4, it was determined that the busway ratings must be 1,600 amp.

Feeder Circuit Breakers

The current rating of the feeder circuit breakers should generally be about the same size as the busways. The 1,600-amp circuit breaker was selected as a standard for all 12 feeder circuit breaker applications.

Main Circuit Breakers and Transformers

The total equivalent continuous current calculated for all the welder loads shown in Fig. 2 is 1,180 amp. Because of the double parallel loops, a 600-amp rating for each transformer should be satisfactory to satisfy the thermal requirements. Later, however, it will be found that 2,[illegible] kva transformers are required becaus[illegible] voltage drop. A 3,000-amp main cir[illegible] breaker was actually selected as a sta[illegible] ard.

Demand and Diversity Factors

In this example, no demand factor [illegible] diversity factor have been used, other t[illegible] those provided for in the method [illegible] determining the equivalent continu[illegible] load for the groups of welders. [illegible] processes become automated to a gre[illegible] degree, it becomes increasingly danger[illegible] to use arbitrary demand factors. If [illegible] known that additional welders have b[illegible] installed in order to keep the product[illegible] line moving at its normal rate with so[illegible] equipment down for repairs, then [illegible] appropriate demand factor can be appli[illegible] Also, if it is known that the average d[illegible] ing-weld current of all the welders c[illegible] nected to one busway, for example, is [illegible] than that for the smaller groups [illegible] welders connected to one busway pl[illegible] then an appropriate diversity factor [illegible] be used.

Summary

This now establishes a thermal rat[illegible] for all elements of the system. It [illegible] important, however, not to forget [illegible] make the second calculation to consi[illegible] the voltage-drop requirement or total d[illegible] ing-weld requirement of all elemen[illegible] These calculations will require so[illegible] of the system elements to be increas[illegible] in ampere rating.

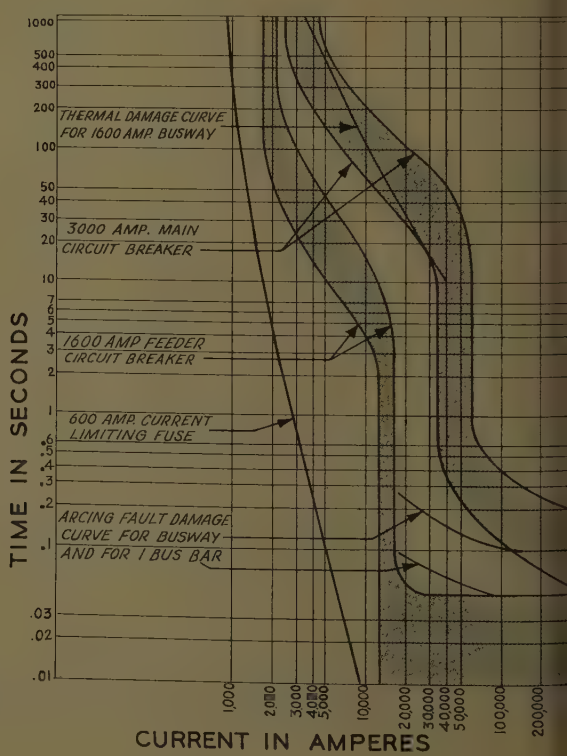


Fig. 4. Time-current curves for protective devices and damage curves for the 1,600-amp busway

Table II. Per-Cent Voltage-Drop Values on 480-Volt Base

System Component	Per-Cent Voltage Drop Per Ampere					
	3-Phase			1-Phase		
	0.3 PF	0.5 PF	0.7 PF	0.3 PF	0.5 PF	0.7 PF
000-kva transformer	0.00065	0.00060	0.00051	0.00075	0.00068	0.00059
500-kva transformer	0.00227	0.00218	0.00185	0.00262	0.00246	0.00214
	Per-Cent Voltage Drop Per Ampere Per 100 Ft					
MCM cable	0.0000015	0.0000017	0.0000019	0.0000017	0.0000019	0.000020
LVDP5 busway	0.00024	0.00029	0.00032	0.00026	0.00031	0.00035
	Per-Cent Voltage Drop Per Ampere					
mbined—1,000 ft of 50-MCM cable and ,000-kva transformer	0.002285	0.002147	0.001869	0.002637	0.002479	0.002160

lution of Problem: Voltage Drop or Total During-Weld Requirement

ses in Busway Plugs

For each group of welding transformers, own in Fig. 2, the ampere ratings of e fuses previously selected to satisfy e thermal requirements should be re-ecked as follows: From the time-current rves of the fuses, determine the time quired for the fuse to blow for the dur-g-weld current indicated in Table I. is important to make this check, to ake sure that the fuse will not blow for is magnitude of during-weld current in e 15 cycles that the during-weld current ws. To allow a little safety factor, it ould be well to increase the size of the se if, in any case, the fuse at this magni-de of current blows in less than, say, cycles.

As an example of this calculation, for e 200-amp fuse previously calculated to tisfy the thermal requirement for 6 lding transformers, the blowing time the fuse for the during-weld current 866 amp is about $7\frac{1}{2}$ seconds. Ob-ously, there would be no danger of the se blowing at this magnitude of current the during-weld time, which is only 15 cles. Note that the probability curves e not used in this calculation because all lding transformers in any one of these ups are fired simultaneously. Since s 200-amp fuse satisfies both the ther-l requirement and the total during-ld current, it can now be recommended the proper fuse to use to protect this up of 6 welding transformers.

eder Circuit Breakers

For the feeder circuit breaker protecting e busway loop 1–4 which has 13 groups welders, it is determined from Fig. 3 at the total number of welders firing ultaneously (bottom scale) for a duty cycle of 0.0277 is 4.3. Before entering the curves in Fig. 3, the number of welders (13) may be reduced by the demand factor, if one has been established. It is dangerous to use a diversity factor in determining the total during-weld current to such a small number (4.3) of welders unless the probability of the average during-weld current for these 4.3 welders exceeding the average for the total group of 13 is checked and found to be less than the probability used in Fig. 3.

$4.3 \times 1{,}732 = 7{,}450$ amp. With the known locations of the large welding presses around the loop, not more than 65% of this total current (about 4,900 amp) will pass through one feeder breaker. The total of 7,450 amp, however, would require almost 10 seconds to trip the 1,600-amp feeder circuit breakers. Certainly, there would be no danger of them tripping in 15 cycles.

Main Circuit Breakers

For the total of 17 welding groups connected to both loops of busways, each having a duty cycle of 0.0277, the total number of welders firing simultaneously from Fig. 3 is 4.7 (use bottom scale.)

Assuming that these groups have 12 transformers each, the total during-weld current would be $4.7 \times 1{,}732$ amp, or 8,150 amp. Of this total current, only about one half would pass through each feeder breaker. Examining the time-current curves of the breaker, as shown in Fig. 4, it is apparent that there would be no danger of the 2,500-amp or 3,000-amp main breakers tripping in 15 or 30 cycles for this value of slightly over 4,000 amp. Thus, either the 2,500-amp or 3,000-amp circuit breakers could be used for the six places in this system where main breakers are required.

Voltage Drop of Busway Loop 1-4

Referring to probability curves in Fig. 3, the number of simultaneous welds to be used for voltage-drop calculations (upper scale) for a total of 13 welders in this loop (for the duty cycle of 0.0277) is 3.3. Since the maximum welder group containing about 12 welding transformers draws 1,732 amp, the total maximum during-weld current is $3.3 \times 1{,}732 = 5{,}720$ amp.

Now, assuming that these 3.3 welders which happen to be firing together are located out near the mid-point of the loop, there would be 2,860 amp flowing in each half of the loop.

From Table II, the per-cent voltage drop of the 1,600-amp type *LVDP5* busway selected for this application is 0.00032% per ampere per 100 feet with the entire load at the end of the run. Thus the per-cent voltage drop in this busway is $0.00032 \times 2{,}860 \times 4 = 3.66$. This appears very satisfactory but, of course, the other elements of the system must be considered.

Voltage Drop in Transformer

For the total of 17 welding groups each with a duty cycle of 0.0277, the number of simultaneous welds from Fig. 3 (top scale) is 3.7. The total maximum during-weld current is (from Table I)) 2,166 for the largest group (with 14 welders) $+2.7 \times 1{,}732 = 6{,}846$.

Of this total current, approximately half would come from each transformer. From Table II the voltage drop in the cable and 2,000-kva transformer previously selected for this application is 0.00187% per ampere at 0.7 power factor. Therefore, the voltage drop for 3,423 amp is 6.40%.

Total Voltage Drop

Then the total voltage drop in the busways and unit substations is 6.4% in the 2,000-kva transformer and cable + 3.66% in the busway, or 10%.

It now remains to consider the voltage drop in the 10,000-kva 138-kv-to-13.8-kv transformer which feeds power to three of these double-ended unit substations.

If the load on the other two substations is assumed to be the same as on this one, there would be a total for 51 welding groups; see Fig. 2. From Fig. 3 there would be 6.2 simultaneous welds. $6.2 \times 1{,}732 = 10{,}800$ amp. From Table II this transformer has a voltage drop of 0.00051% per amp or 5.5% for 10,800 amp.

Then the total voltage drop is $10 + 5.5 = 15.5\%$. This appears satisfactory, especially in view of the fact that the other two unit substations are not expected to be as heavily loaded as the one shown in Fig. 2. But changes were required in

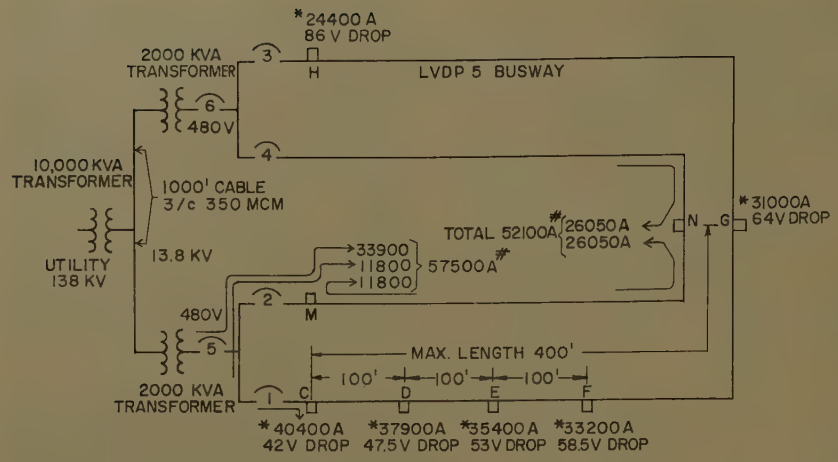


Fig. 5. Calculated values of short-circuit current and division of current in parallel paths

* Calculated for radial busway connected to one transformer with breakers No. 2 and No. 3 open

\# Calculations for points M and N are for all breakers closed

the ampere ratings of the system elements as selected to satisfy the thermal requirements originally, thus illustrating the importance of making voltage drop calculations.

Selection of Protective Devices to Achieve Selectivity and Maximum Protection for Equipment and Personnel

Selectivity

One of the best ways to obtain selectivity between the feeder circuit breaker protecting the busways in these double parallel loops and the main circuit breakers would be to use the selective system employing breakers, such as the General Electric type *AK-1*. In this system, the main breaker is equipped with a delay of a few cycles, just sufficient to allow the feeder breakers to trip without tripping the main breaker. It is believed that it is very important to keep this delay at an absolute minimum, because if an arcing fault should occur in the system between the main breaker and the feeder breaker, it is desirable for the main breaker to trip as quickly as possible.

In the specific sizes of breakers used in this double-loop system, it appears that selectivity can be achieved by using standard circuit breakers with the usual inverse time delay for overloads in addition to the instantaneous trip device.

In the overload range, Fig. 4 shows that the 1,600-amp feeder breakers will trip for all overloads without tripping the 3,000-amp main breakers.

To consider the selectivity in the short-circuit region, Fig. 5 shows that, with both loops operating in parallel, the maximum possible short-circuit current is 57,500 amp, at busway plug M, near the input ends of any of the busways and about 52,000 at the farthermost point N, which is at the mid-point of one loop. This assumes an 800-foot loop. The short-circuit value at any intermediate point around the loop would be something between these two values. For the 57,500-amp short circuit at the input end the feeder breaker, with instantaneous trip device set at 10 times, will be tripped by its instantaneous trip device by the 45,700-amp current.

These short-circuit current values as calculated were based on the 1.25 factor. Actually the X/R ratio in this Twinsburg plant varies from 2.7 to 4.4. A more realistic average factor is 1.1 instead of 1.25 for a 3-phase short circuit. In case of a single-phase short circuit, the factor would ordinarily be higher. In these low-voltage systems, however, a single-phase short-circuit value is reduced to about 86% of the 3-phase value. Therefore, it appears that for circuit breaker–co-ordination studies the factor of 1.1 should be used for all calculations of maximum possible short-circuit currents, either 3-phase or single-phase. Thus, the maximum short-circuit value at point M is $57{,}500 \times 1.1/1.25 = 51{,}000$ amp and, for point N, 46,000 amp.

The 1,600-amp feeder breaker would trip instantly, in 3 cycles, since, from Fig. 5, the current passing through it is approximately (33,900+11,800) times the ratio $1.1/1.25 = 45{,}700 \times 0.88 = 40{,}000$ amp.

The current passing through the main 3,000 amp is $0.88 \times 33{,}900 = 30{,}000$ amp. Thus, if its instantaneous trip device is set at 12 times or 36,000 amp it will not trip. Calculations indicate that none of the other breakers will trip. It appears then that satisfactory selectivity has been achieved providing a high degree of protection for the maximum possible values of short-circuit current. This is an interesting advantage of the double-loop system. In the radial system, with a 3,000-amp main breaker and 1,600-amp feeder breakers, selectivity cannot be achieved for values of short-circuit current higher than 36,000 amp ($12 \times 3{,}000$) because the instantaneous trip devices of both the feeder breaker and main breaker will be actuated in the first half cycle.

Arcing Faults

In case of an arcing fault, the current would be reduced by the arc resistance to a value estimated as approximately 70% of these values on a 480-volt system, or about 40,000 amp for the fault near the input end and 36,000 amp at the farthermost point. These values are symmetrical values to take care of the minimum possible condition.

It is considered to be extremely important to trip the feeder breakers instantly, in case of an arcing fault, to reduce to a minimum the damage done by the arcing and the danger to personnel. Since the magnitude of arcing fault current for a fault at the midpoint of one run is only 36,000 and only about half this value, or 18,000 amp, would come from each direction, it then appears important to set the instantaneous trip device of the 1,600-amp feeder breakers not higher than 10 times, so they will trip instantaneously for any current over 16,000 amp.

Another point that must be checked carefully is an arcing fault current near the input end of the busway at busway plug M, in Fig. 5. When the fault first occurs at this point with all breakers closed, the short-circuit arcing fault current magnitude is about 40,000 amp symmetrical ($0.70 \times 57{,}500$) with about 31,000 amp passing through the feeder breaker No. 2. With this breaker set at 10 times, its instantaneous trip device will function. Then Fig. 6 shows the magnitude of current that would flow from the other direction through feeder

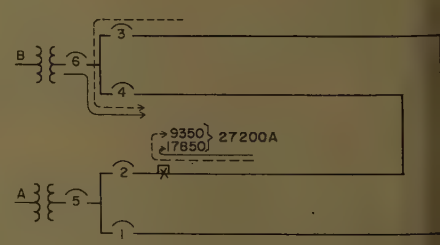


Fig. 6. Short-circuit currents after breaker No. 2 trips

eaker No. 4. The maximum value is ,200 amp but, for an arcing fault, this uld be reduced to 70% or about 19,000 p symmetrical. Fortunately, this lue will trip the breaker No. 4 if its stantaneous trip device is set not higher an 10 times. Perhaps 8 times would a safer setting.

High-speed tripping was given careful tention in the design, because it is own that considerable damage is done arcing faults in industrial plants. It the authors' opinion that the most im-rtant single action that can be taken to duce this damage is to select with great re the instantaneous trip setting of all rcuit breakers, to make sure that they ill trip to interrupt the current quickly—not more than 4 or 5 cycles—for an cing fault occurring at any point in the stem.

Fig. 4 shows the time-current curves d also the instantaneous trip points the 1,600-amp and 3,000-amp circuit eakers.

Superimposed on this curve is the ermal damage curve for the busway. his curve shows how long the busway can rry given amounts of current without ceeding its temperature limitation of 55 (degrees centigrade) rise. Also, on this rve are shown the arcing damage curves r the more massive aluminum conductor the joints where all bus bars of like olarity are connected together and also r individual bars.

Since the bus bars of the *LVDP* bus-ay are taped there is no danger of arcing cept at the bare spots at the joints or at e plug-in openings. In this type of bus-ay the busway plugs make contact with ree of the multibar interlaced arrange-ent of bars. The fuses in the busway ugs protect the branch circuit, and the dividual bus bars for a short circuit the load side of the fuses. For an ternal arcing fault on these individual s bars, however, the fuse, not being the circuit, cannot provide any protec-on. For this reason it is desirable to t the instantaneous trip device on the eder circuit breaker low enough to pro-ct the individual bars.

Referring to the curve showing the cing fault damage for one bar of the sway, this curve shows about how ng a given amount of current can be lowed to flow through an arcing fault rning on one bar in the busway with-t destroying more than 50% of the nductor material. Reading from the rve, the bus bar could stand 20,000 p for about 5 cycles and 100,000 amp r about 3 cycles. The newer circuit eakers that trip in the range from 2 to 4 cycles appear to be fast enough provided the instantaneous trip device is set low enough.

Some engineers may question the advisability of attempting to protect from arcing faults, believing that regardless of what is done replacement may be necessary anyway. It is the authors' opinion, however, that low-voltage arcing faults are so destructive and dangerous that far greater care should be exercised in selecting protective devices so that the damage and danger will be reduced to a minimum.

Arcing Faults in Switchboards

In this study both the maximum possible values of short-circuit current and also the minimum arcing fault currents have been considered for short circuits at all points in the busways. No consideration has been given to a short circuit in the unit substation switchboard between the main breaker and the feeder breaker. A short circuit on the line side of one of the feeder circuit breakers, for example, would be approximately the same as at point M, Fig. 5, or 40,000 amp minimum for an arcing fault. Only about 24,000 amp minimum would pass through the main breaker, whereas its instantaneous trip device is set at 36,000 amp. Under this condition the main breaker would trip after considerable delay, but the instantaneous trip device would not function. Considerable damage would be done to the bus bars in the switchboard.

To provide high-speed protection for arcing faults in the switchboard without losing selectivity, it appears necessary to use induction disc relays to trip the main breakers after a delay of, say, 8 cycles. This would provide high-speed tripping yet the 8 cycles would allow sufficient time for the feeder breaker to operate in 3 cycles to interrupt current to any short circuit on its load side without tripping the main breaker. Another equally good arrangement would be the use of the short time delay in the selective system, such as on the type *AK-1* circuit breakers manufactured by the General Electrical Company.

Emergency Condition

In an emergency it is possible that the feeder breakers 1 and 4 may be closed leaving the feeder breakers 2 and 3 open; refer to Fig. 5. The entire length of 800 feet of busway would be fed from 1 transformer. Under this condition the short-circuit current at point H, the extreme end of one run, would be 24,400 amp maximum with only about 17,000 amp in case of an arcing fault at the end. With the instantaneous trip device of circuit breaker 1 set at 10 times, it should trip to give protection for this arcing fault. This is a borderline case, however, and there is a chance that the breaker might not trip on the instantaneous trip device. A setting of 8 times might be required.

Another possible solution would be to place a disconnect switch at the midpoint of both loops. Then the maximum length of busway would be only 400 feet, and the short-circuit current would be a little higher, so the feeder breaker should trip satisfactorily even for an arcing fault.

Co-ordination with Fuses

In selecting the point at which the instantaneous trip devices are to trip, it is important, of course, to consider how they co-ordinate with the fuses in the busway plugs. On a system having this much short-circuit current available, only current-limiting fuses in the busway plugs should be used for two reasons. First, the fuses must have adequate interrupting ability. Second, it is highly desirable for these fuses to have a low value of I^2t to provide greater safety for arcing fault currents and provide a reduction in the forces on the smaller circuits which they protect.

Most current-limiting fuses blow so quickly on short circuit that there would be no danger of tripping the feeder circuit breakers before they blow, but it is important to actually check the curves to make sure that complete selectivity is achieved.

New Plug-In-Type Low-Voltage-Drop Busway

For the double parallel loops of busways feeding power to the heavy welding presses, a new type of plug-in busway was selected. This busway employs the interlaced arrangement of bus bars with general details of construction similar to the standard General Electric type *LVD* feeder busways. Fig. 7 shows a fused busway plug mounted on one of the 10 openings available in a 10-foot length of the new busway known as G-E type *LVDP*. These busways are available in a range of ratings from 600 to 4,000 amp with either copper or aluminum bus bars. Any 10-foot length of this *LVDP* busway is interchangeable with a 10-foot length of standard *LVD* feeder busway.

Type *LVDP* busway rated 1,600 amp with aluminum bus bars was selected for this Chrysler distribution system to feed the heavy welder loads.

This busway with 4 bus bars in parallel per pole is capable of withstanding the

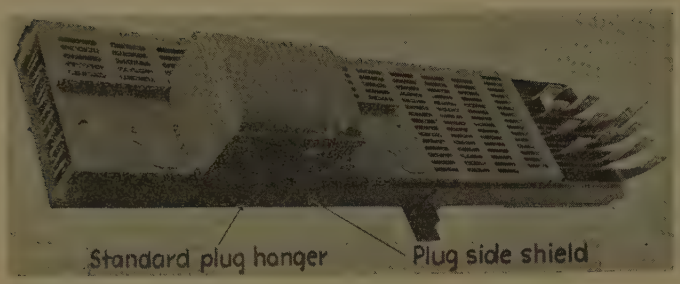


Fig. 7. Type LVDP plug-in high-capacity busway

forces of a 100,000-amp short circuit. The 4 bus bars in parallel of each pole are connected together at every joint between 10-foot sections. The 100,000-amp short-circuit current divides among the 4 bus bars with about 25,000 amp per bar. This, together with the fact that the bus bars are effectively supported on insulators 12 inches apart, explains its high short-circuit strength.

In the spans between insulator supports of any busway, the forces during short circuit cause the bars to swing back and forth. In order to reduce the effect of this movement of the bus bars on the plug-in stabs of the busway plugs, the 10 plug-in openings in each section of busway are positioned directly adjacent to the bus bar supports. Also the stabs and stab insulators are designed with sufficient flexibility to permit some movement. Based on tests on other similar devices, no damage is expected in either the busway or busway plugs for a 100,000-amp short circuit on the load side of current-limiting fuses mounted in the busway plugs.

Among the most important features of these plug-in busways are extremely low values of resistance, reactance, and voltage drop. The voltage-drop value is roughly half that of the conventional plug-in busways.

Assuming a current of 1,600 amp at a power factor of 0.70 on the 1,600-amp plug-in busway selected for this welder distribution system, the voltage drop is only 1.22 volts per 100 feet with loads distributed along the run. With the entire 1,600-amp load at the the end, the voltage drop is 2.44 volts per 100 feet.

Use of Series Capacitors to Improve the Voltage Regulation

It is known that series capacitors can be used to improve voltage regulation on systems with welder loads. The a-c network analysis test Case No. 15 proved that a remarkable reduction in voltage drop could be achieved on the Chrysler system by using series capacitors.

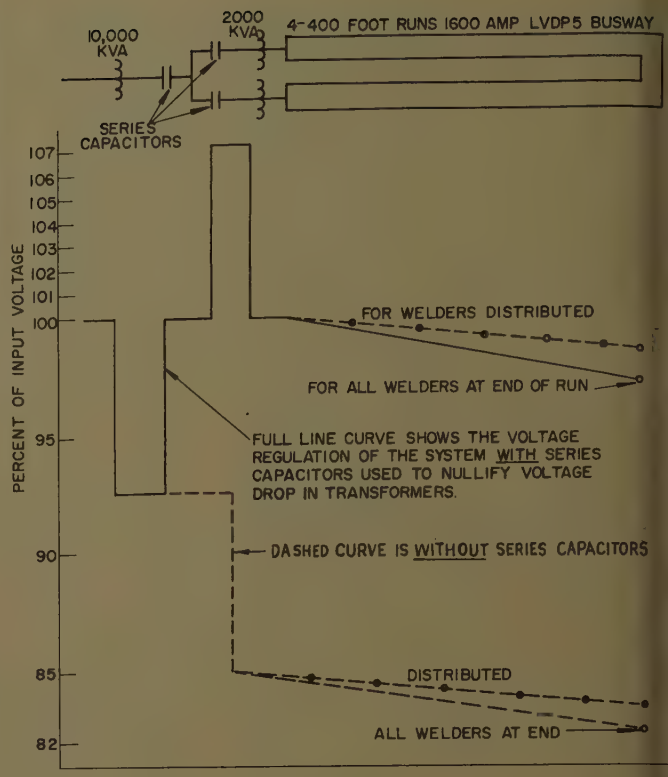


Fig. 8. Voltage levels on Chrysler welder system with and without series capacitors

The graph in Fig. 8 shows calculated values of voltage drops or levels, first with and then without the use of series capacitors to nullify the voltage drop in the transformers.

After some preliminary trials it was determined that all the series capacitors can be grouped together near the 10,000-kva main transformer. It is important, however, to break the total bank into units, one for the 10,000-kva transformer placed in series with it, and one for each 2,000-kva transformer, being careful to place each one in the individual branch circuit in series with 1 2,000-kva transformer.

The reason is that the series capacitor is unusual in that it can exactly nullify the voltage drop in a transformer or busway, and the voltage drop will be nullified for all values of load current from zero to maximum, provided that all the current which passes through the transformer or busway also passes through the capacitor, and vice versa. Thus, if only one capacitor bank were used and placed on the load side of the 10,000-kva transformer, it could be adjusted to nullify the voltage drop in all transformers if all are fully loaded. But it is obvious that any decrease of current in any one or more of the branch transformers would cause a voltage drop in all the other branches in which the current had not changed. This is because the drop in current passing through the capacitor would make a reduction in the voltage rise in the part of the capacitor which was being used t nullify the drop in the other branches.

Possibility of Leading Power Factor

Actually, only a small amount of c pacitor kva is required to nullify the vol age drop in transformers or busways, sinc the voltage drop in these devices is onl 5 or 10% of the total voltage. Thus would appear that there would be n possibility of having a leading powe factor. Theoretically, there is some po sibility, however, because when the sy tem is very lightly loaded to say, onl 5 or 10% with high power-factor load, would then be possible to have a leadin power factor if the capacitor selected larger than necessary to nullify the i ductive kva in the transformers or bu ways. This is of little importance how ever.

Series Capacitors for Busways

If all the load is connected at the en of the feeder, a capacitor can be selecte so that the voltage at the end remai unchanged for all values of load from n load to full load. It is evident that th voltage level varies along the length of th run. Therefore, for busways such plug-in busways where the loads are di tributed along the length of the run, it not possible to select series capacitors th will completely nullify the voltage drop all points along the busway. Thus the appears little advantage in using a seri

Table III. Electrical Parameters

tem Component	Ohms	Per-Cent Ohms, 666-Kva Base	Per-Unit Ohms, 20,000-Kva Base
y 138 kv			
00,000 kva S.C.)	1.42 +j7.4	0.005+j0.026	0.0015+j0.0078
0,000 initially)	5.9 +j30.8	0.022+j0.109	0.0066+j0.0327
former, 10,000 kva	0.104 +j1.52	0.036+j0.53	0.0108+j0.159
, 1,000 ft	0.040 +j0.032	0.014+j0.011	0.0042+j0.0033
former, 2,000 kva	0.00126+j0.0062	0.28 +j1.80	0.084 +j0.54
ay *LVDP5*			
100 ft	0.00081+j0.00044	0.233+j0.1275	0.070 +j0.0383
400 ft	0.00324+j0.00176	0.933+j0.51	0.28 +j0.153

citor for plug-in busways. One be used to hold the voltage level con- t at about the one-third point in the but there will be a voltage rise at the t end and a drop at the output end l to about half the normal drop.

ne Chrysler system was designed to ide satisfactory voltage regulation out using capacitors with the in- ion of considering their use at a later when production increases beyond a ain level.

Network Analyzer Tests

ne electric distribution system feeding welder load at the Ohio Stamping t of Chrysler Corporation was ulated by a 3-phase test on the a-c net- k analyzer at the General Electric npany in Schenectady, N.Y.

arameters of the 10,000-kva 138-kv sformer, the 13.8-kv cables, the 2,000- 13.8-kv transformers, and the double s of *LVDP* busways as tabulated in le III were set up on the analyzer.

xteen tests or case studies were con- ed using the network analyzer:

1. Fault at point *A* at center of foot loop of *LVDP5*. Base kva = 00; unit current = 24,060; amp = 100%. l tests with 400-foot runs of *LVDP5* ay; 800-foot loops unless noted other- . On all tests, the base voltage is volts line to line or 277 volts line to ral.

2. Fault at secondary side of one -kva transformer. Base kva = 20,000; current = 24,060; amp = 100%.

3. Apply 1 4,000A 3ϕ load at point 0.30 PF (power factor). Base kva = ; unit current = 3,610; amp = 100%. e values remain the same for all re- ing tests.

4. Apply 1 4,000A 3ϕ load at point d one at point *B* 0.30 PF.

5. Apply 1 4,000A 1ϕ load at point 0 PF.

6. Apply 1 4,000A 3ϕ load at point *A* 1ϕ at point *B* 0.30 PF.

7. Apply 1 8,000A 1ϕ load at point 0 PF.

8. Apply 1 8,000A 1ϕ load at point 0 PF.

Case 9. Apply 1 8,000A 1ϕ load at point *B* 0.70 PF.

Case 10. Connect busway radial with both loops open. Apply 8,000A load 1ϕ 0.3 PF.

Case 11. Apply 1 4,000A 3ϕ load at point *A* and 1 8,000A 1ϕ at *B* 0.3 PF.

Case 12. Repeat Case 11 with length of busway changed from 400 feet to 200 feet.

Case 13. Add 80 foot current limiting busway between the 2 2,000-kva transformers. Apply 1 4,000A 3ϕ 0.3 PF load on 1 transformer. Apply 1 8,000A 1ϕ 0.3 PF load on other transformer.

Case 14. Apply fault at 10-12-14 on Case 13 with load connected.

Case 15. Insert series capacitor on secondary of 10,000-kva transformer, repeating Case 11.

Case 16. Repeat Case 11 with 1,500-kva transformers.

The results of the network analyzer tests were in close agreement with calculated values of short-circuit current and voltage drop. Valuable information was obtained from the tests to aid in making decisions on the design of future systems. As an example, test number 15 indicated that series capacitors can be used to reduce, or even nullify, the voltage drops in transformers and feeder busways. The tests show the advantage of the double loop system of busways in reducing the voltage drop, providing additional current-carrying ability, because of the parallel runs, or providing greater flexibility in firing the welders or in the location around the loops; and even in providing better selectivity between feeder and main circuit breakers.

Although it is not the purpose of this paper to present in detail an analysis of these tests, Table IV shows the voltage-drop values for certain selected loads placed at the mid-points of the busway loops.

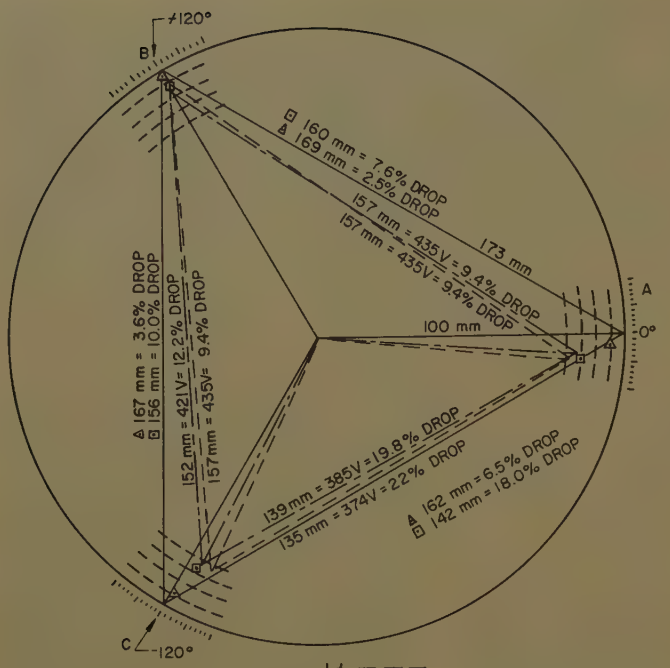
The tests using combinations of heavy single-phase welder load current superimposed on the busway system at the same time it is carrying balanced 3-phase currents are interesting. It proves that the voltage drop under these conditions can be calculated with sufficient accuracy by the arithmetical addition of the voltage drop determined by the busway manufacturer's literature for the 3-phase current and that for the single-phase current. Obviously, the 3-phase voltage drop tables must be used to determine the drop due to the balanced 3-phase current, and the single-phase tables for the single-phase current. Complications may arise in some types of busways, making it necessary for the manufacturer to provide a third set of tables showing the voltage drop values for a single-phase current on 1 phase of a 3-phase busway. In the *LVDP* busway used on this Chrysler job this voltage drop value happens to be the same as the values for single-phase currents on a single-phase busway.

Illustrative of the data or information obtained from the a-c network analyzer tests is the 3-phase vector diagram shown in Fig. 9 for Case 11. On this test the single-phase current of 8,000 amp, in addition to the 3-phase current of 4,000 amp, exceeds the maximum values anticipated in actual production, but it makes a good example since the voltage-drop

Table IV. Voltage-Drop Values from A-C Network Analyzer Test

	Per-Cent Voltage Drop				Total Per-Cent Voltage Drop at Load			
	10,000-Kva Transformer	Cable and 2,000-Kva Transformer	Busway to Point A	Busway to Point B	Point A	Point B	Load at Point A*	Load at Point B*
3	2.1	3.9	2.0		8.0		3,651	
4	4.2	7.5	1.8	1.8	13.5	13.5	3,426	3,423
5	Not recorded	6.8		2.7	9.5			3,602
6	4.2	8.0	2.6	2.7	14.5	14.9	3,523	3,390
7	5.4	7.5		4.6		17.5		6,600
8	4.8	8.1		4.5		17.0		6,570
9	4.2	6.3		5.8		16.3		6,570
10	6.5	12.2		8.0		26.7		5,890
11	6.5	11.5	2.2	4.0	19.8	22.0	3,406	6,210
12	6.9	11.5	1.3	2.6	19.5	21.0	3,436	6,280
13	6.3	11.2	4.0	8.0	19.8	25.5	3,320	5,920
15	−19.0	16.3	2.1	3.9	−0.6	1.2	3,946	7,800
16	6.0	15.1	1.9	3.2	23.0	24.7	3,330	6,000

* See Fig. 2.



	PERCENT VOLTAGE DROP	
	4000A 3φ	8000A 1φ
AB	9.4%	9.4%
BC	12.2%	9.4%
CA	19.8%	22.0%

Fig. 9. Three-phase vector diagram showing voltage levels determined in a-c network analyzer test 11

values are exaggerated (unusually large). The single-phase load of 8,000 amp is connected to phase conductors A and C.

The result is 6.5% drop in the 10,000-kva transformer, 18% drop in the 2,000-kva transformer, and about 4% drop in the busway, making a total voltage drop in phase AC of 22%.

Conclusions

It is believed that the method used, both in the example in this paper and for the Twinsburg system, is very effective in selecting the proper system elements and it takes many factors into consideration.

1. It provides system elements adequate to satisfy the thermal requirements, voltage-drop requirements, short-circuit requirements with protective devices that will co-ordinate properly, providing selectivity and high-speed tripping for both the maximum values of short-circuit current and the minimum values of arcing fault current. This provides greatly increased protection for the equipment since the damage done by arcing faults is a function of the arcing time. More important, it provides greater safety to electricians working with the equipment.

2. The use of the probability curves permits the selection of transformers, cables, or busways with smaller ratings than would be required otherwise.

3. The double parallel loops of plug-in busways provide parallel paths for the welding current, effectively increasing the busway's current-carrying ability, and reducing the voltage drop in both the busways and transformers. It also makes it easier to obtain selectivity between the main and feeder circuit breakers.

4. The new plug-in busways used to feed the heavy welding departments are ideally suited for this application. They have adequate short-circuit strength, and the voltage drop is reduced to almost half that of conventional plug-in busways.

5. Series capacitors can be used to nullify partially or completely the voltage drop in transformers and also in feeder busways.

6. Vector algebraic calculations of short-circuit current values and current division around the parallel paths in the busway loops (Appendix I) or an a-c network analyzer study is necessary in order to properly select system elements.

Appendix I. Calculations

Short-Circuit Current at Mid-Point G in Fig. 5, Using Impedances on 666-Kva Base from Table III

The 666-kva base was chosen for simplifying the final calculation of the short-circuit current at 480 volts. The final asymmetrical (1.25) value $=100\times1.25\times666\times1{,}000/(\sqrt{3}\times480\times Z)=100{,}000/Z$ in whi Z is the per-cent impedance on 666-kva ba Because of symmetry, no current will fl in inner busway loop, M-N-4.

Impedance	Per Cent
1,000 ft of 15-kv cable	0.014+j0.01
2,000-kva transformer	0.280+j1.80
400 ft of busway	0.933+j0.51
Total circuit through C to G or H to G	1.227+j2.32
Combined imp. of two circuits in parallel to point G	0.613+j1.16
Add the utility	0.022+j0.10
Plus the 10,000-kva transformer	0.036+j0.53
Total to point G	0.671+j1.79

Short-circuit current at $G=100{,}000/1.92=52,$ amp asymmetrical (1.25 factor).

Short-Circuit Current at Point M Fig. 5 Assuming 800-Foot Busw Loops

Impedance of path A through cable a 2,000-kva transformer to M
$=Z_A=0.294+j1.811=1.84$
$Z_A{}^2=3.34$

Likewise, impedance of path B throu cable and transformer to H
$=0.294+j1.811$

Impedance of 2 800-foot runs of busway parallel from H to M
$=0.932+j0.51$

Total impedance of path B to point M
$=Z_B=1.226+j2.321=2.62$
$Z_B{}^2=6.88$

The admittance $Y=1/Z=g+jb=r/Z$ $j\,x/Z^2$
$Y_A=0.088+j0.528=0.544$
$(0.544/0.924=59$
$Y_B=0.178+j0.337=0.380$
$(0.380/0.924=41$
Y Total $=0.266+j0.865=0.907$
$Z^2=(1/0.907)^2=1.22$
$Z=gZ^2+jbZ^2=0.324+j1.05$

Now adding the impedance of the utility	$=0.022+j0.10$
And the impedance of 10,000-kva transformers	$=0.036+j0.53$
Total impedance to point M	$=0.382+j1.69$ $=1.7$
Short-circuit current at point M	$=100{,}000/1.74$ $=57{,}500$ a

59% or 33,900 amp flows through path to point M

41% or 23,600 amp flows through path with 11,800 amp flowing through e busway loop as indicated in Fig. 5

Short-Circuit Current at Point M

Assume:

1. Two 10,000-kva transformers parallel.

2. Each busway loop 500 feet long.

Using the same method as in the ceding calculation, the total short-cir current at point $M=71{,}400$ amp 40,000 amp flowing from the nearest tr former and 15,700 amp flowing thro each of the two busway loops to point M

ort-Circuit Current at Point G with 2 10,000-Kva Transformers in Parallel

Combined impedance of the two paths in rallel as calculated in Appendix I.

$$Z = 0.613 + j1.16$$

Utility $= 0.022 + j0.109$

0,000-kva transformers in parallel $= 0.018 + j0.265$

Total impedance to $G = 0.653 + j1.534 = 1.64$

$$I_{sc} = 100{,}000/1.64 = 61{,}000 \text{ amp}$$

ltage Drop and Short-Circuit Current at Points C, D, E, F, G, and H

nditions:

oncentrated load of 4,000 amp at 50% power factor at point *H*

rcuit breaker at point *H* is open

efer to Table II for voltage-drop data for busway and transformer

oltage drop in transformer A at 4,000 amp and 5.5% impedance $= 0.00218 \times 4{,}000 \times 480/100 = 42$ volts

oltage drop in 100-foot length of type *LVDP5* busway at 4,000 amp $= 0.00029 \times 4{,}000 \times 480/100 = 5.5$ volts

urther calculations will be found in Table V

Table V. Calculations for Various Voltage-Drop Values

Voltage Drop Values Volts to Point		Short-Circuit Calculations
		$0.022 + j0.109$ ..... utility
		$0.036 + j0.503$ ..... 10,000-kva transformer
		$0.014 + j0.011$ ..... cable
		$0.280 + j1.8$ ..... 2,000-kva transformer
42	*C*	$0.352 + j2.45 = 2.47\%$ $I_{sc} = 40{,}400$ A
		$0.233 + j0.127$ ..... busway: 100 ft
42 + 5.5 = 47.5	*D*	$0.585 + j2.58 = 2.64\%$ $I_{sc} = 37{,}900$
		$0.233 + j0.127$ ..... busway: 100 ft
42 + 11 = 53	*E*	$0.816 + j2.70 = 2.82$ $I_{sc} = 35{,}400$
		$0.233 + j0.127$ ..... busway: 100 ft
42 + 16½ = 58.5	*F*	$1.05 + j2.83 = 3.08$ $I_{sc} = 33{,}200$
		$0.233 + j0.127$ ..... busway: 100 ft
42 + 22 = 64	*G*	$1.28 + j2.96 = 3.22$ $I_{sc} = 31{,}000$
		$0.93 + j0.51$ ..... busway: 400 ft
42 + 44 = 86	*H*	$2.22 + j3.47 = 4.10$ $I_{sc} = 24{,}400$

breaker 4 to the fault at *M*. By calculation similar to Appendix I, this reduced value is 27,000 amp, as shown in Fig. 6.

Calculations such as indicated for different points around the loops made it possible to set the instantaneous trip devices of the circuit breakers to achieve selectivity and a high degree of protection for the equipment and personnel.

It is of interest to record that an actual arcing fault experienced at about point *D*, Fig. 5, resulted in feeder breakers 1 and 3 tripping instantly without tripping breakers 2, 4, 5, or 6, exactly as predicted by these calculations. Although some damage was done to the busway at that point, not a single bus bar was burned off more than half way, exactly as predicted. The 10-foot length of busway could have been repaired temporarily by retaping 3 or 4 bars at the point where the arcing occurred and repairing the steel cover of the housing. The burned notches in the bar would make no appreciable reduction in their current-carrying ability.

Of even greater importance, however, the danger to personnel was greatly reduced by the extremely fast tripping of both feeder breakers.

Appendix II. Sequence of Circuit Breaker Tripping

Short-circuit current calculated at point *M* in Appendix I with both loops of *LVDP* busway operating in parallel was 57,500 amp. The 45,700 amp flowing through breaker 2 as indicated in Fig. 5 will trip breaker 2 instantly. Then, a reduced value of short-circuit current flows through

References

1. Power Supply for Resistance Welding Machines. *AIEE Special Publication S-45*, Apr. 1952.

2. Equivalent Continuous Current for Groups of Variable Loads, W. K. Boice. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1938–41.

3. Probabilities of Interference Between Resistance Welders, W. K. Boice. *Ibid.*, pt. I, 1951, pp. 775–83.

Discussion

I. J. Donnelly (New York Telephone Company, Albany, N. Y.): The authors re to be congratulated on the excellence f their paper and on the design and on horoughness of all phases of their investigaon. Of the many outstanding design eatures involved there is only one upon hich I would elect to comment, main econdary breaker versus feeder breaker o-ordination philosophy. This paper is he first major work to come to my attention here the particular case treatment of rcing faults as well as maximum short rcuits has been thoroughly and accurately elineated. With the thought that much ood may come from widespread study of he paper, I would encourage the authors o expand their views on this matter to nclude the more general case.

In my background training, I was led o believe that, depending on X/R ratio, or medium-voltage 8-cycle breakers a ultiplier of 1.6 includes the d-c offset, nd that this factor was conventional. urthermore, although low-voltage breakers see" the same 1.6 offset, breaker manuacturers say their low-voltage equipment ill perform satisfactorily against 3-phase short-circuit values computed from a multiplier of 1.25 times calculated symmetrical current. Also, when one of the phase devices is considered, a multiplying factor for total current at the first half cycle of 1.5 is used.

Based on multiplying factor versus time curves for various values of X/R published by one manufacturer, the multiplying factor to be applied at the first half cycle for X/R values of 2.7 to 4.4 would seem to be of the order of 1.1 to 1.3, or to average 1.20. Nevertheless, in the ordinary calculation, subtransient reactance is employed to compute symmetrical short-circuit current, and the 1.25 recommended factor is used to compute interrupting or momentary duty. This is ordinarily ratioed up to 1.5 and applied for co-ordination studies with the thought that if co-ordination is achieved at maximum fault, all other values will also co-ordinate.

If the objective of a selective system is to localize disturbance as much as possible, it would appear that the authors may wish to offer some further explanation of their reasons for electing settings based on a factor of 1.1. On casual observation, it seems under some conditions and on systems in general, backup breakers might be brought into operation unnecessarily. In the particular case, with the feeder breaker set at 10 times, for instantaneous, the question, of course, has no material bearing. In many other cases it may be of importance, however, and it is for these I would believe the authors may wish to amplify their reasoning.

The authors have made an excellent analysis of arcing faults. Their study assures adequate treatment for feeders and they have carefully and forthrightly pointed out the difficulty in protecting against faults in switchgear buswork. I am aware of the more or less general practice of establishing selectivity at the price of moving the instantaneous trip setting of the main secondary breaker out of range, or of not furnishing the instantaneous trip in the first place.

I am sure the authors' precept, "to provide as much selectivity as possible, to provide high-speed tripping for both the maximum possible values of short-circuit current and also for the lower values of arcing fault current," is optimum. However, with the technique employed, they say, "The main secondary breaker would trip after considerable delay, the instantaneous trip device would not function, and considerable damage would be done to the bus bars in the switchboard."

I am acquainted with a situation in a large commercial building in the network area of a large city. A short circuit in the main secondary buswork, calculated to approximate about 6 to 8,000 rms amp symmetrical, was permitted by the main secondary breaker to persist. An electrical fire resulted, some 400 employees evacuated the building, the power company had to shut down the downtown network to stop the fire, a very expensive temporary and a new permanent installation had to be made. The switchboard installation had been placed in service only three months prior to the fire.

Now I realize the desirability of preventing shutdown of all or a large part of certain chemical or manufacturing plant power services for faults on one feeder. Furthermore, in these particular establishments, feeder faults may be a fairly common experience. Also, a large monetary loss might result, in certain special cases, before a main secondary breaker could be reclosed to re-establish power to the extent possible, after a feeder fault had tripped a main breaker.

There is some question in my mind, however, as to whether enlightened industrial building or commercial building managements would waive adequate and certain protection on switchboard buswork merely to insure full selectivity of a system. It has been argued by some that secondary bus faults are rare and thus may be discounted. When they do occur, however, they can be tragic. If such were to happen, it might cause complete shutdown for an indefinite period. Conversely, from the undue emphasis placed upon it, one would be led to believe that feeder faults are, in general, an everyday experience. In current industrial and commercial building practice, sufficient care is devoted to design and installation that I think this is no longer the case, if it were ever such.

Because of the serious nature of switchboard bus faults, I believe the majority of managements, by and large, would prefer shutdown of a main secondary bus on a certain few feeder faults, to underwriting the risk of a sustained low amperage bus fault. The relative incidence of failure on switchboard bus versus feeders appears immaterial, since both are rare. What would seem to matter most is that in case of bus fault, the current is interrupted fast enough to prevent serious damage to the switchboard.

S. G. Forbes (General Electric Company, Richland, Wash.): This paper is a very interesting account of a large installation of production welders in what must be a rather completely automated plant. The technical aspects of the problems involved, the methods used to arrive at solutions, and the justification of those solutions are presented with a marked tone of condescension. The general impression indicates this should be an article in a periodical such as *Iron Age, Manufacturing*, or one of the other semitechnical papers commercially published to give general coverage to manufacturing design problems and incidental publicity to some companies' products.

The paper is not detailed enough in scope, method, or justification to present a truly technical problem as a *Transactsons* paper. The use of proprietary nomenclature for components leaves a reader unfamiliar with the particular catalog designations used completely in the dark as to the capabilities of the equipment chosen. Such a reader would be unable to verify the suitability of the items referred to as being used for particular purposes.

The problems discussed and the solutions used are neither unique nor original. The presentation tends toward the promotional rather than technical discussion of the solution of one customer's problems by a vendor happily having made a sale.

W. K. Boice (General Electric Company, Erie, Pa.): The authors have presented a very interesting review of a problem illustrating the practical application of probability data.

Provision for further expansion or changes of the load has been made by the use of conservatism in the selection of load conditions assumed for calculations. Additional elements of conservatism are introduced by the method of calculation. The authors' calculations obtained the total voltage drop by addition of drops calculated for each of the main circuit elements, such as busways and transformers.

This procedure neglects diversity which can occur among voltage drops in these different circuit elements. The more precise method of calculation suggested in reference 1 of the paper results in calculated voltage drop about 20% less than the drop calculated by the authors. An additional safety margin is introduced because the total drops are sufficient to perceptibily reduce the current drawn by all the welders. As explained in reference 3 of the paper, this further reduces the drop by about 12%.

The factors mentioned tend to reduce the number of welds spoiled by low voltage. However, the paper does not indicate any specific allowance for variations in the utility supply voltage.

It would be constructive if the authors could provide information obtained from actual experience with the distribution system as installed, possibly comparing calculated results with the actual proportion of welds spoiled by low voltage.

T. J. Higgins (University of Wisconsin, Madison, Wis.): I have read this paper with very considerable interest. It comprises a clearly written and carefully documented account of a logical procedure for the rational design of a heavy-duty bus distribution system for welding loads. The conjunction of analytic calculation, enabling a choice of parameters and structure, satisfying prescribed thermal and electrical requirements, and check-out of these on the a-c network analyzer provides a gratifying correlation and confirmation of the flexibility and soundness of the design as a whole.

L. E. Fisher, R. W. Dailey: The authors wish to thank the discussers for their interest and comments.

In response to Mr. Donnelly, the 1.1 factor resulted from the X/R ratio of the system involved. It is recognized however, that the maximum offset in one pole of th main breaker has a ratio greater than 1. and that the current in any one pole ma trip the breaker. Some provision for th was made by setting the instantaneous tri higher than the calculations required.

Although the calculation using the 1. average value appears to be an oversimpl fication, the detailed problem encountere by breaker application engineers is com plicated. The 1.1 factor is only a benc mark. To make certain that a breaker wi trip, an additional factor must be considered Then, to make certain that a breaker wi not trip, another additional factor is neces sary. This factor is related to the maximun offset that can occur in any one pole. Sinc this paper was written, more accurate dat are being developed to aid in the selectio of the trip setting to take into considerationo this maximum offset.

To prevent the main breaker from trip ping in many cases it is believed that it i necessary to introduce a short time dela in the tripping of the main breaker. Be cause of the destructive nature of arcing i is believed that this delay should be no much more than 8 cycles.

With the protective devices now nor mally used in low-voltage distribution sys tems it is often difficult to achieve complete protection for arcing faults. Furthermore this protection, if given any considerationo whatsoever, is often compromised in orde to obtain selectivity.

As the study of this problem of arcing faults in low-voltage equipment continues it becomes clear that a ground fault tripping device is often necessary to supplement the direct acting trips on circuit breakers.

For example, in many large 460Y/265 volt spot network systems feeding com mercial buildings, even where the short circuit level is above 100,000 amps, it has been determined that the line-to-neutra arcing fault current is so low that fuses cannot blow in time to prevent serious arcing damage. Two of the three fuses may blow quickly, after which the arcing may con tinue from the one remaining energized phase conductor to the switchboard or bus way housing for 10 minutes or more, causing complete destruction.

This type of extensive damage could be prevented by the use of a circuit breaker equipped with a ground fault relay which will sense current flow in the connecting strap between the switchboard or busway housing and the transformer neutral.

With circuit breakers as used in thi Chrysler Plant all three poles would be tripped instantly after line-to-line arcing is established. Nevertheless, on any 3ϕ 4 wire grounded system the arcing often originates from one phase to the housing Therefore, it now appears that a significant improvement in arcing fault protection can be made by using a ground fault relay, be cause it could sense even low values of line to-housing arcing fault currents and trip the breaker faster than it could be tripped by the direct acting trip devices.

Perhaps the above discussion relating to Mr. Donnelly's discussion will in a way answer Mr. Forbes' criticism. There are many problems involved in low-voltage distribution systems that have not yet been solved to the satisfaction of the engineers Often the highly theoretical treatment of a narrow phase of the total problem, although

nely important, is overshadowed by problems such as selectivity and arc- hat have not been treated so precisely. ther examples, the practical problem taining an accurate measurement of g-weld time and weld pressure has in- d much study by the Chrysler engineers their solutions would appear to be ny of a future AIEE paper.

r. Boice's comments are appreciated ially in view of the fact that the mate- on probability used in the paper was ed from Mr. Boice's papers, as in- ed in references 2 and 3. We agree completely with Mr. Boice that as the voltage drops there is a corresponding reduction in the current drawn by all the welders drawing current at that time and, therefore, the net resulting voltage drop is reduced by about 12%. Therefore, a 15% voltage drop would be reduced to 13.2%.

Diversity was considered in adding the voltage drop of the busway to that of the transformer. Although the addition appears to be arithmetical it is actually vectorial since the per-cent voltage drops of all elements in Table II take into consideration the R/Z factor of the element and the load power factor. Also, the during-weld currents were obtained independently for each element, based on the probability on the total number of welders taking current through that element.

Realizing that this paper covers a multitude of problems, whereas the ideal AIEE paper is intended to cover a narrow subject in detail, it is the authors' belief that such a paper is necessary to present the broad systems concept and point out the interrelation of these narrow subjects in the overall picture.

Instrument Servomechanisms with Backlash, Coulomb Friction, and Stiction

M. P. PASTEL
MEMBER AIEE

G. J. THALER
MEMBER AIEE

N INSTRUMENT servomechanism is normally thought of as a ll positioning device with low power sumption. A typical block diagram is wn in Fig. 1. A gear ratio of 1 to 1 used in this paper for convenience. e load is shown as an inertia with lomb friction but no viscous friction. is further assumed that $J_L=0$. This ght correspond to the common case which the output member is just the er arm of a potentiometer. When re is backlash in the gear train the owing sequence of events is normal. e motor drives (with backlash taken), reducing the error: as the motor ocity decreases the load velocity decreases, and continuous mechan- l contact is maintained because $J_L \cong 0$, d the coluomb friction C provides equate drag on the output. When the tor velocity becomes zero, the load ocity also becomes zero. If the or is not zero, the motor reverses but load does not move because of the cklash. Thus, the error does not nge and the motor is driven open- p with a constant driving torque, voltage, until the backlash is taken . The load is then "picked up" by the drive and both the output velocity and the error rate jump from zero to the instantaneous drive shaft velocity. This sequence is usually repeated a number of times.

It is well known that the effect of backlash is destabilizing. The net result of the condition just described is certainly a step response with extended duration of oscillation, and may result in a limit cycle, or in a static error. Which of these alternatives exists depends on the viscous damping ratio, ζ, as is shown in the stability criterion derived in this paper. If coulomb friction is added to the system it provides additional damping and helps to stabilize the system. The coulomb friction can be added at either the motor shaft or the load shaft with appreciably different performance characteristics. When stiction is present the normal effect on the step response is just a decrease in static accuracy as is shown by phase plane studies.

Definitions and Nomenclature

Stiction is the dry friction torque which determines the threshold; i.e., the torque at which motion is initiated. Coulomb friction is the dry friction torque effective during running conditions.

C = coulomb friction torque
D = constant evaluated from initial conditions
E = error $= \theta_R - \theta_C$, when gears are contacting
E_N = error variable when gears are separated
f_M = motor friction torque per unit velocity
J_M = polar moment of inertia at motor
K = system torque constant
t = time
t_Δ = time required to take up backlash
Δ = amount of backlash
ζ = damping ratio
λ_{max} = angle of tangent to backlash recombination lines
θ_C = angular position of load
$\dot{\theta}_C, \ddot{\theta}_C$ = load velocity and acceleration
θ_R = angular position of input
$\theta_M, \dot{\theta}_M, \ddot{\theta}_M$ = position, velocity, and acceleration of motor
ρ = radius in polar co-ordinate representation
ψ = angle in polar co-ordinate representation
ψ_T = angle of backlash separation line on polar plane
ψ_{max} = angle of tangent to backlash separation line on polar plane

Backlash Dividing Lines on the Phase Plane

Assume that a step displacement input is applied to the servomechanism, and assume that the gears are in contact in the proper direction so that the backlash is not effective and the system starts to respond as a linear system. The basic equations are

$$J_M\ddot{\theta}_C + f_M\dot{\theta}_C = KE \quad (1)$$

$$E = \theta_R - \theta_C \quad (2)$$

In the phase plane of Fig. 2, these equations give the first segment of the trajectory from M to N. At point N, $\dot{E} = 0$ and $\dot{\theta}_C = 0$. The motor reverses, accelerates, and increases its velocity and displacement but the output does not move and the error does not change until the backlash is taken up. Since the error remains constant at $E = E_N$ during this

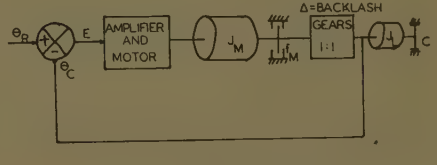


Fig. 1. Block diagram

er 60-118, recommended by the AIEE Feedk Control Systems Committee and approved by AIEE Technical Operations Department for sentation at the AIEE Winter General Meeting, w York, N. Y., January 31–February 5, 1960. nuscript submitted March 12, 1959; made ilable for printing December 8, 1959.

P. PASTEL and G. J. THALER are both with U. S. val Postgraduate School, Monterey, Calif.

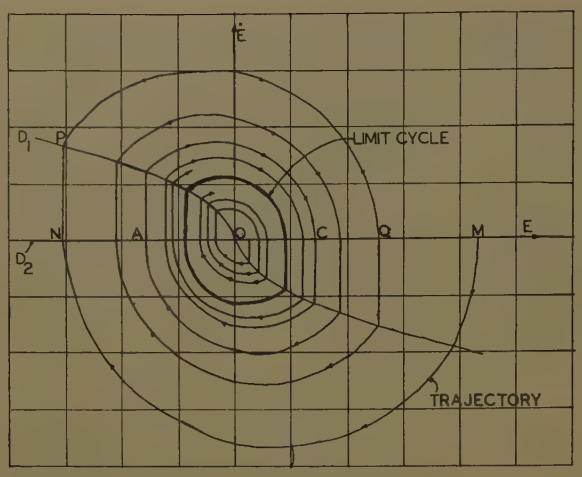


Fig. 2 (left). Phase trajectories of an instrument servomechanism with backlash. $D_1 \leftrightharpoons$ dividing line for recombination, $D_2 \leftrightharpoons$ dividing line for separation

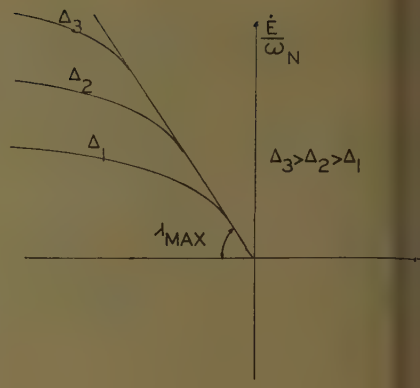


Fig. 3 (right). Family of backlash dividing lines with common tangent at the origin. $\lambda_{max} \leftrightharpoons \tan^{-1} 1/2\zeta$

period, the system is operating open loop according to the law

$$J_M\ddot{\theta}_M + f_M\dot{\theta}_M = KE_N \tag{3}$$

from which

$$\theta_M = -KE_N\left[\frac{t}{f} + \frac{J_M}{f^2}\left(e^{-\frac{ft}{J_M}} - 1\right)\right]$$

$$= \frac{-E_N}{4\zeta^2}\left(-1 + 2\zeta\omega_n t + e^{-2\zeta\omega_n t}\right) \tag{4}$$

where

$$\zeta = f/2\sqrt{KJ_M} \text{ and } \omega_n = \sqrt{K/J_M}$$

Also

$$\dot{\theta}_M = -\frac{\omega_n E_N}{2\zeta}\left(1 - e^{-2\zeta\omega_n t}\right) \tag{5}$$

Note that at point N the operation of the system changes from the laws defined by equations 1 and 2 to the laws defined by equations 4 and 5. There are many points at which this transition occurs, and they form a locus which may be called a "dividing line," since it divides the phase plane into several regions in which different equations apply. This particular transition always occurs when $\dot{E}$ goes through zero, so the E-axis in the phase plane of Fig. 2 is a backlash dividing line. Equations 4 and 5 apply during take-up of the backlash; that is, until the displacement described by equation 4 is such that $\theta_M = \Delta$ where Δ is the backlash in radians. During this period E and $\dot{E}$ do not change so the tracing point in Fig. 2 remains at N. Solving equation 4 for t when $\theta_M = \Delta$, the velocity $\dot{\theta}_M$ is determined by substituting this value of t in equation 5. Thus, after the time interval, t_Δ, the backlash is taken up, the output assumes the same velocity as that of the motor shaft, and the tracing point on the $\dot{E}$-versus-E-plane jumps from N to P. From P to Q the trajectory is again governed by equations 1 and 2, at Q, equations 4 and 5 take over, etc. Note that there must exist a locus of points such as P, at which the describing equations change as indicated. This locus is a second backlash dividing line. In general, the location of this line is not apparent from inspection and must be computed point by point if the phase trajectory is to be constructed. Such a line has been computed for Fig. 2. Note that equation 5 can be rewritten

$$\frac{\dot{\theta}_M/\omega_n}{E_N} = \frac{-1}{2\zeta}\left(1 - e^{-\zeta\omega_n t_\Delta}\right) \tag{6}$$

and for $t_\Delta \to \infty$ the slope of the dividing line is $-1/2\zeta$. Thus, for various values of backlash, Δ, the backlash lines form a family of curves which branch out of a common tangent through the origins of the phase plane as shown in Fig. 3.

Limit Cycle on the Phase Plane Stability Criterion

The phase trajectory in Fig. 2 shows a limit cycle. The construction of the trajectory may start at a large displacement, in which case the trajectory spirals toward the origin; or the construction may start near the origin, in which case the trajectory spirals outward, away from the origin. In either case, the final result is the same closed path, which represents a continued, fixed amplitude oscillation or limit cycle. It is apparent that this limit cycle is due to the presence of backlash. In this section a stability criterion is developed which shows that the existence of a limit cycle in a second-order system depends not only on the presence of backlash, but also on the amount of damping, ζ. The dividing lines of Fig. 2 have polar symmetry, and therefore a limit cycle must exist if, for instance, $OA = OC$. If $OA > OC$, the amplitude of oscillation decreases; and if $OA < OC$ the amplitude increases. These conditions, determined from the phase plane by inspection, give rise to a method for determining the existence of a limit cycle for this type of system. For the second-order system mathematical relationship can be dete mined and presented in curve form. detailed derivation of the criterion given in the Appendix; in outline form t procedure is to convert the equation the phase trajectory into polar co-ord nates, obtaining

$$\rho = D^2 e^{-\frac{\zeta\psi}{\sqrt{1-\zeta^2}}}$$

and transferring the backlash dividi lines to the new phase plane as shown Fig. 4. The system is unstable; i.e., t oscillations increase in amplitude, i Fig. 4,

$$\rho_1 \cos\psi_1 \leqq \rho_2 \cos\psi_T$$

Combining these equations, the stabili criterion becomes

$$\cos\psi_1 \leqq \sqrt{1-\zeta^2}\, e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}(\pi - \psi_1 - \psi_T)}$$

but this is not useful because ψ_1 is not pr cisely defined. However, near the orig ψ_1 approaches ψ_{max} and if the inequali exists for $\psi_1 = \psi_{max}$ then it is certain th a limit cycle exists. Substituting t proper functions for ψ_1:

$$1 \leqq \frac{1}{2\zeta}\, e^{-\frac{\zeta}{\sqrt{1-\eta^2}}\left(\pi - \tan^{-1}\frac{1-2\zeta^2}{2\zeta\sqrt{1-\zeta^2}} - \tan^{-1}\frac{\zeta}{\sqrt{1-\zeta^2}}\right)}$$

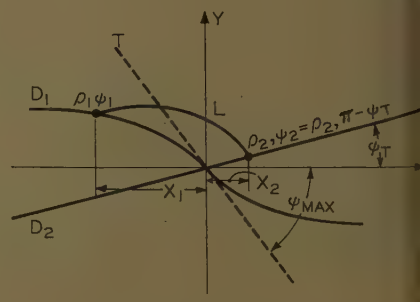
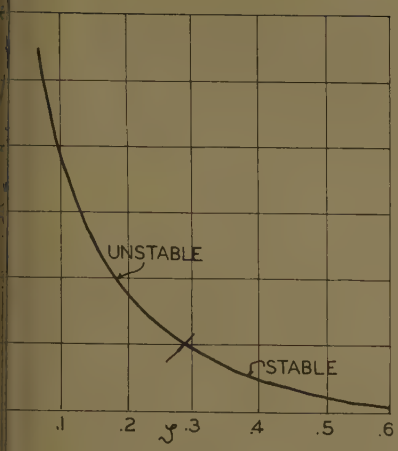


Fig. 4. Definition of the phase trajectory co-ordinates.

$X_1 \leftrightharpoons \rho_1 \cos\psi_1$, $X_2 = O_2 \cos\psi_T$,.
$\psi_T \leftrightharpoons \tan^{-2}\zeta/\sqrt{1-\zeta^2}$
T $\leftrightharpoons$ tangent to dividing line at origin.
L $\leftrightharpoons$ segment of logarithmic spiral



5. Graph of stability criterion for backlash

$$e^{-a(\pi-\tan^{-1} b-\tan^{-1} a)}$$

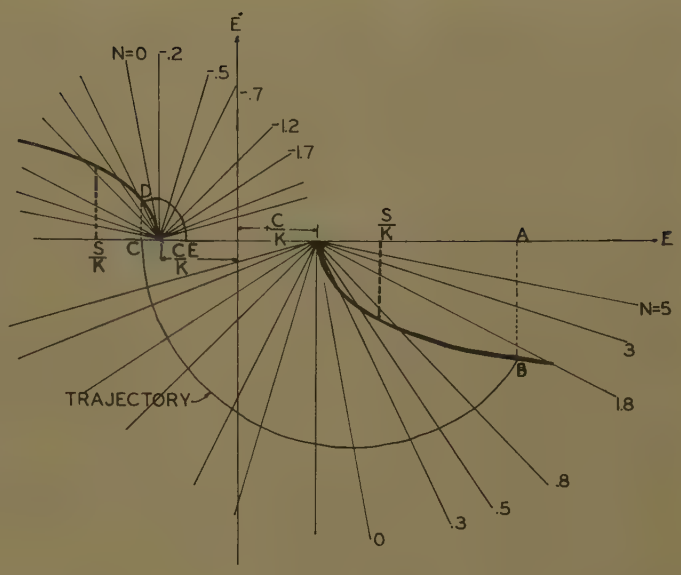
$$2\zeta$$

$$/\sqrt{1-\zeta^2}$$

$$(1-2\zeta^2)/2\zeta\sqrt{1-\zeta^2}$$

lot of equation 10 is shown in Fig. 5, it is seen that any second-order sys- with backlash is unstable (has a limit e) if $\zeta \leqq 0.29$. Note that this is due the presence of backlash, but inde- dent of the amount of backlash. If a t cycle exists, the amplitude of the t cycle depends on the amount of klash, but the existence of the limit le is determined by ζ. Further insight his condition may be obtained by in- ction of Fig. 3; existence of a limit le depends on the angle λ_{max} which is unction of ζ. If ζ is large and the le is small, no limit cycle exists. When limit cycle does exist, the phase tra- ory converges to the knee of the back- dividing line, and the location of this e depends on the amount of backlash, ch determines the amplitude of the t cycle.

Fig. 6. Effect of coulomb friction and stiction when added at the motor shaft



The Effect of Coulomb Friction

If coulomb friction is deliberately added to a system which already contains backlash, the damping effect of the coulomb friction may be used to eliminate the limit cycle. This friction may be added to the drive or to the load. In either case the system equations are affected, but with different results, as will be shown.

Coulomb Friction at the Motor Shaft

For this case equation 1 becomes:

$$J_M\ddot{\theta}_C+f_M\dot{\theta}_C+C \text{ sign } \dot{\theta}_C=KE \qquad (11)$$

As shown previously[1] this causes a split and translation of the focal point. In like manner equation 3 becomes:

$$J_M\ddot{\theta}_M+f_M\dot{\theta}_M+C \text{ sign } \dot{\theta}_M=KE_N \qquad (12)$$

and the backlash dividing line is also split and translated as shown in Fig. 6. The trajectory in each half plane is governed by the isoclines from the focus indicated, and it is apparent that convergence toward steady state must be more rapid due to the presence of C. If a trajectory terminates on the E-axis between C/K and $-C/K$, the value of E_N is insufficient to initiate motion and a static error exists. Fig. 7 shows a computed trajectory for a second-order system illustrating this effect. When discontinuous stiction[2] is present, the linear phase trajectories are not affected, but the backlash dividing line is slightly altered as shown in Fig. 6, and the width of the dead zone is increased so that a larger static error is possible. For example, if there is coulomb friction but no stiction a trajectory such as $ABCDE$ in Fig. 6 is possible, with a rather small static error. With stiction present the same size disturbance produces trajectory ABC, with a larger static error.

Coulomb Friction at the Load Shaft

For this case equation 11 still applies, but when the system stops and the motor reverses, the coulomb friction does not hold the motor shaft, so equation 3 applies. As a result the coulomb friction splits and translates the focal point providing the same isocline configuration shown in Fig. 6, but the backlash dividing

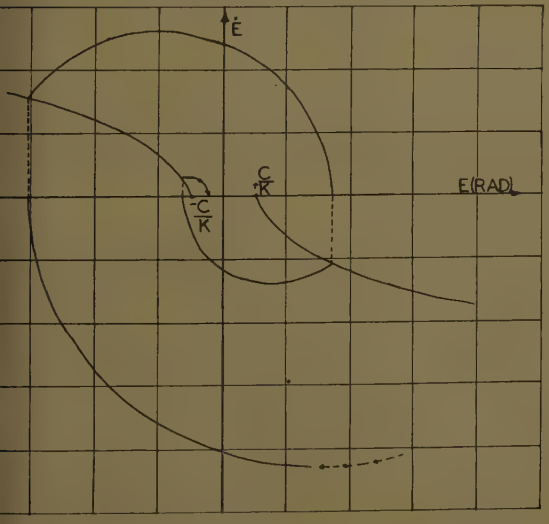


Fig. 7 (left). Stabilizing effect of coulomb friction at the motor shaft

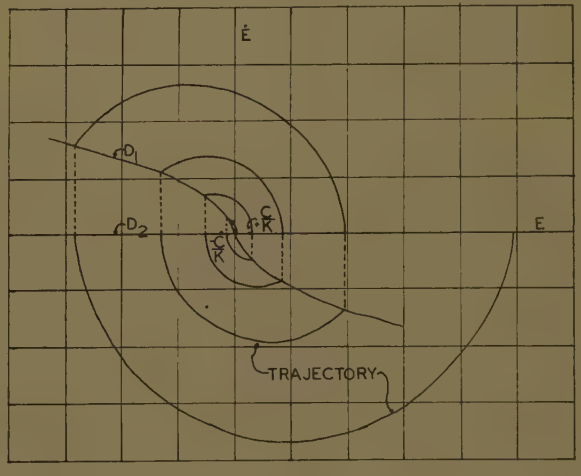


Fig. 8 (right). Effect of coulomb friction at the load shaft

line is not affected by the coulomb friction and still goes through the origin. Fig. 8 shows a computed plot indicating that the coulomb friction stabilizes the system and reduces the error to zero. It is apparent that large amounts of coulomb friction are needed to stabilize a system which is lightly damped initially, and has a large amount of backlash, but this can be done without impairing the static accuracy if the coulomb friction is added on the output side of the backlash. It appears that adding still more coulomb friction, while it may increase the rise time and the settling time, will also provide a well damped system with no static error. This suggests the interesting possibility that static error may be eliminated by operating the servo mechanism at very high gain, deliberately introducing backlash at the output measuring device, and damping the oscillations with coulomb friction on the measurement side of the backlash. It is possible to use too much coulomb friction, however. Note that the phase trajectory which passes through the origin of the phase plane must lie on or inside the backlash recombination line, or a static error obtains. There is also a unidirectional threshold value when coulomb friction is used in this fashion. When the system comes to rest after a disturbance, the backlash is all taken up in one direction. If the next disturbance is in the proper direction to go through the backlash there is no threshold, but if the disturbance is not so directed the magnitude of the disturbance must be great enough to overcome the coulomb torque, or motion is not initiated.

Conclusions

Backlash in an instrument servomechanism is destabilizing, but for the case of a low inertia load a limit cycle does not obtain unless ζ, the damping ratio of the linear system, is less than 0.29. The addition of coulomb friction stabilizes the step response. If the coulomb friction is added on the drive side of the backlash the static error is increased, but if added on the load side of the backlash the coulomb friction does not affect the static error. Stiction on the drive side of the backlash increases the static error, but does not if on the load side.

Appendix. Derivation of Stability Criterion for Backlash Only

If equations 1 and 2 are combined and nondimensionalized forming a single equation with error, E, as the independent variable, this equation can be manipulated and integrated to define the linear phase trajectory on the phase plane as:

$$\dot{E}^2+2\zeta\omega_n E\dot{E}+\omega_n^2E^2 = D^2e^{\frac{2\zeta}{\sqrt{1-\zeta^2}}\tan^{-1}\frac{E+\zeta\omega_nE}{\omega_nE\sqrt{1-\zeta^2}}}$$

Using the linear transformation

$$Y=\dot{E}+\zeta\omega_nE,\ X=\omega_nE\sqrt{1-\zeta^2}$$

$$X^2+Y^2=D^2e^{\frac{2\zeta}{\sqrt{1-\zeta^2}}\tan^{-1}\frac{Y}{X}}$$

Letting $Y=\rho\ \sin\ \psi$, $X=\rho\ \cos\ \psi$, $\psi=\tan^{-1}Y/X$

$$\rho=De^{-\frac{\zeta\psi}{\sqrt{1-\zeta^2}}}$$

The backlash dividing lines are located on the $X-Y$ plane by noting that

$$\frac{Y}{X}=\frac{\dot{E}+\zeta\omega_nE}{\omega_nE\sqrt{1-\zeta^2}}$$

but the dividing line for separation is the E-axis, for which $\dot{E}=0$, and for this line

$$\frac{Y}{X}=\frac{\zeta}{\sqrt{1-\zeta^2}}=\tan\psi_T$$

The backlash dividing line for recombination is not readily determined, but the tangent to this curve λ_{max} on Fig. 3, is easily transferred:

$$\frac{Y}{X}=\frac{\dot{E}/\omega_n}{E}\frac{1}{\sqrt{1-\zeta^2}}+\frac{\zeta}{\sqrt{1-\zeta^2}}$$

and for $\lambda_{max}\ \dot{E}/E\ \omega_n=-1/2\zeta$ so

$$\frac{Y}{X}=\frac{2\zeta^2-1}{2\zeta\sqrt{1-\zeta^2}}=\tan\psi_{max}$$

which may be expressed as an acute angle

$$\tan\ \psi_{max}=\frac{1-2\zeta^2}{2\zeta\sqrt{1-\zeta^2}}$$

In Fig. 4 the backlash line for recombinati is sketched, not calculated, but the bas condition for increasing oscillations is

$$\rho_1\cos\psi_1=\rho_2\cos\psi_T$$

Assuming that the angles are measur from the negative real axis in Fig. 4, a that ρ_1 and Z_1 are known:

$$D=\rho_1e^{\frac{\zeta\psi_1}{\sqrt{1-\zeta^2}}}$$

$$\rho_2=De^{-\frac{\zeta\psi_2}{\sqrt{1-\zeta^2}}}=\rho_1e^{-\frac{\zeta(\psi_2-\psi_1)}{\sqrt{1-\zeta^2}}}$$

$$\psi_2=\pi-\psi_T$$

Substituting,

$$\cos\ \psi_1\leqq\cos\ \psi_Te^{\frac{-\zeta(\pi-\psi_T-\psi_1)}{\sqrt{1-\zeta^2}}}$$

$$\leqq\frac{1}{\sqrt{1-\zeta^2}}\ e^{\frac{-\zeta(\pi-\psi_T-\psi_1)}{\sqrt{1-\zeta^2}}}$$

The angle ψ_1 may be chosen arbitrari and by inspection of Fig. 4 it is appare that if the required inequality exists must be most pronounced when the poi $(\rho_1,\ \psi_1)$ is located near the origin on t tangent line, in which case $\psi_1=\psi_m$ Substituting this value for ψ_1:

$$1\leqq\frac{1}{2\zeta}\ e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\left(\pi-\tan^{-1}\frac{1-2\zeta^2}{2\zeta\sqrt{1-\zeta^2}}\right.}$$

$$\left.-\tan^{-1}\frac{\zeta}{\sqrt{1-\zeta^2}}\right)$$

References

1. Stability Criteria for Instrument Servomechanisms with Coulomb Friction and Stiction, M. P. Pastel, G. J. Thaler. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. Nov. 1959, pp. 294–97.

2. Backlash in a Velocity Lag Servomechanism, N. B. Nichols. *Ibid.*, vol. 72, 1953 (1954 section), pp. 462–67.

Discussion

C. Concordia (General Electric Company, Schenectady, N. Y.): We noted with particular interest the authors' remark that coulomb friction could be used to eliminate the limit cycle of a system containing backlash. This may be true in specific instances, but we should like to call attention to the fact that it is not so as a generality. Indeed, coulomb friction is in many ways analogous to backlash and can in some cases produce effects that are practically indistinguishable from it. This can be shown either from the describing-function point of view or from differential-analyzer studies. For example, this is pointed out in the appendix of reference 1 of this discussion.

With respect to stiction, it should be remarked that this also can result in self-excited oscillations (as shown in reference 1 of this discussion).

We have made these comments primarily in order to point out a possible danger in extrapolating the results of these studies of second-order systems to the general higher-order case, and not to reflect in any way on the validity of the authors' work within the range of his study.

Reference

1. Effect of Prime-Mover Speed Governor Characteristics on Power-System Frequency Variations and Tie-Line Power Swings, Concordia, S. B. Crary, E. E. Parker. *AI Transactions*, vol. 60, Jan.-Dec. 1941, pp. 559 and 734.

M. P. Pastel and G. J. Thaler: We wo like to thank Mr. Concordia for his co ments, and state that we are in agreem with him.

To extend his comments still farther, should be noted that our results are valid for all second-order systems, but o for that class of second-order systems hav negligible inertia on the load side of backlash.

When some (or all) of the system ine is on the load side of the backlash

equations become more compli- conservation of momentum must be ered in the analysis, and construction phase portrait[2] of a second order becomes difficult; its general inter- on still more difficult. The order of ty increases manyfold with only slight increases in the order of the differential equation. To date, work in these areas has been far from comprehensive.

It is possible that beneficial effects may be obtained by using coulomb friction damping, but this has not been substantiated.

REFERENCES

1. ANALYSIS OF SERVOMECHANISMS WITH GEARS HAVING BACKLASH, M. P. Pastel. *Master's Thesis*, University of Washington, Seattle, Wash., 1948.

2. DIVIDING LINES FOR BACKLASH IN THE PHASE PLANE, W. J. Lutkenhouse. *Master's Thesis*, U. S. Naval Postgraduate School, Monterey, Calif., 1959.

ut-Output Cross-Correlation Functions for Some Memory-Type Nonlinear Systems with Gaussian Inputs

HAROLD R. LELAND
NONMEMBER AIEE

previous paper[1] the author de- ibed the application of the method si-linearization (that of Booton[2]) to etermination of input-output cross- ation functions for relay servo- anisms. Experimental verification validity of this analysis was pre- l in the form of analog-computer ation results for several specific con- tions. The present work is an sion of the earlier effort, presenting esults for simulated systems with ry-type nonlinearities. The theoret- nalysis of the memory-type system complished by combining quasi- ization with a well-known analog- uter technique for generating resis-type nonlinearities by placing emory nonlinearities within closed ops.

oximation of the Hysteresis aracteristic

e type of hysteresis characteristic studied is that shown in Fig. 1. loop is of width $2b$ and the input- ut relations indicated are for differ- mplitude periodic signals. This type steresis is often provided in analog uters by use of the equipment shown . 2. Here a dead-space or dead-zone element is cascaded with an integrator and a high-gain element, C_H, and this forward chain is provided with unity negative feedback. The combination gives the input-output characteristic of Fig. 1; it is, however, frequency-sensitive, and when using this configuration in analog computing one must provide sufficiently large C_H to make high-frequency distortion of the hysteresis loop negligible. The dead-zone element of Fig. 2 must have total deadzone of $2b$ to provide a hysteresis loop corresponding to that of Fig. 1.

Using the same procedures as in reference 1, the equivalent gain of the dead-zone element of Fig. 2 is readily derived. Let the input be $y(t)$ and the output $x(t)$. Then

$$x(t)=K_{dz}y(t)+e_1(t) \tag{1}$$

where $e_1(t)$ is the error in representing $x(t)$ approximately by $K_{dz}y(t)$. (K_{dz} is the equivalent gain of the dead-zone element.)

The square of this error, whose mean is to be minimized, is

$$e_1^2(t)=[g(y)-K_{dz}y]^2 \tag{2}$$

where $g(y)$ is the function:

$$\begin{aligned} g(y)&=d(y-b) \text{ for } y>b \\ g(y)&=0 \quad\quad \text{for } -b>y<b \\ g(y)&=d(y+b) \text{ for } y<-b \end{aligned} \tag{3}$$

Assume that the signal $y(t)$ is an element of an ergodic stationary, Gaussian, random process and that the first probability-density function is given by:

$$p_1(y)=\frac{1}{\sqrt{2\pi\beta^2}}\,e^{-\frac{y^2}{2\beta^2}} \tag{4}$$

where β is the rms value of $y(t)$. The mean-square value of $e_1(t)$ is

$$\begin{aligned} \langle e_1(t)\rangle = &\int_{-\infty}^{\infty} g^2(y)p_1(y)dy-2K_{dz}\times \\ &\int_{-\infty}^{\infty} g(y)yp_1(y)dy+K_{dz}^2\times \\ &\int_{-\infty}^{\infty} y^2p_1(y)dy \end{aligned} \tag{5}$$

Simplification of this expression yields:

$$\begin{aligned} \langle e_1(t)\rangle = &K_{dz}^2\beta^2+2d^2b^2\int_b^{\infty} p_1(y)dy+ \\ &4K_{dz}db\int_b^{\infty} yp_1(y)dy- \\ &4db^2\int_b^{\infty} yp_1(y)dy+2d^2\times \\ &\int_b^{\infty} y^2p_1(y)dy\ 4K_{dz}d\int_b^{\infty} y^2p_1(y)dy \end{aligned} \tag{6}$$

Equation 6 is to be minimized with respect to K_{dz}, hence

$$\begin{aligned} \frac{\delta\langle e_1(t)\rangle}{\delta K_{dz}} = &2K_{dz}\beta^2+4db\int_b^{\infty} yp_1(y)dy- \\ &4d\int_b^{\infty} y^2p_1(y)dy=0 \end{aligned} \tag{7}$$

and

$$K_{dz}=d\left[1-E_2\left(\frac{b}{\sqrt{2\beta^2}}\right)\right] \tag{8}$$

where $E_2(\ )$ is the error function, following the definition of reference 3.

The result expressed in equation 8 is plotted as the solid curve in Fig. 3, with K_{dz}/d shown as a function of the ratio β/b. Note that K_{dz}/d is always less than unity.

The approximtion to the hysteresis loop to be used is obtained by replacing the dead-zone block of Fig. 2 with a gain element, K_{dz}, as indicated by dashed lines. The frequency-response function for this loop is

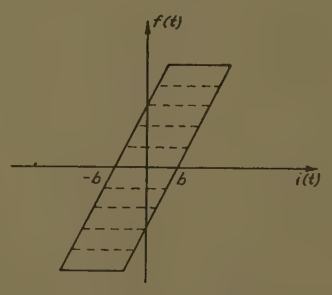


Fig. 1. Hysteresis characteristic

60-115, recommended by the AIEE Feed- Control Systems Committee and approved AIEE Technical Operations Department esentation at the AIEE Winter General g, New York, N. Y., January 31–February 0. Manuscript submitted July 13, 1959; available for printing December 7, 1959.

D R. LELAND is with Cornell Aeronautical tory, Inc., Cornell University, Buffalo,

uthor acknowledges the guidance and as- e of Prof. V. C. Rideout, University of sin, Madison, Wis., in conducting the h upon which this paper is based.

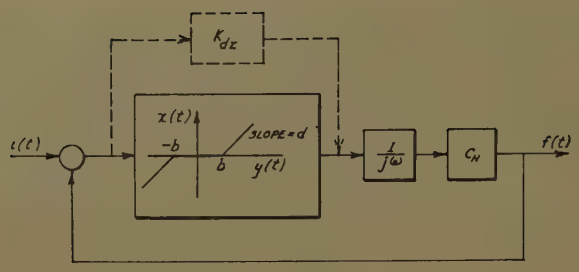


Fig. 2 (left). Hysteresis unit using dead-zone characteristic and integrator

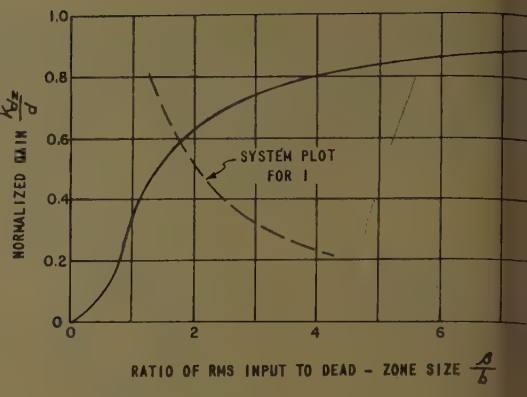


Fig. 3 (right). Equivalent gain of dead-zone element

$$\frac{F}{I}(j\omega)=\frac{K_{dz}C_H}{j\omega+K_{dz}C_H} \tag{9}$$

Thus, the effect of hysteresis of the type being considered is represented in this quasi-linear approximation by a simple lag element. When the hysteresis element is a component of a larger system, the procedure to be used in analysis of the system is as follows. First, the magnitude of C_H is selected. Then the quasi-linearized system is solved for the rms level of the input to the block containing K_{dz} and a curve of rms input level versus K_{dz} is added to Fig. 3. The intersection of this "system plot" and the graph of K_{dz} as a function of β gives the value of K_{dz} to be used in the quasi-linearized system, and with K_{dz} and C_H determined, straightforward, linear analysis completes the determination of the input-output cross correlation function.

Closed-Loop Servomechanisms with Hysteresis: Examples

As a preliminary to the analysis of a system which includes a relay with hysteresis or "holding," two systems including an element with the input-output characteristic of Fig. 1 are analyzed. An element with this characteristic has two important applications. It may be used to represent backlash when, for example, a shaft driven through a loose coupling has no inertia or spring return. It may also be used to provide an approximation to iron-core hysteresis if cascaded with a modified form of sharp limiter.[4] The first of the two systems to be analyzed has a hysteresis element in the forward portion of the control loop, and is designated System 1. System 2 represents a physically significant case in which the feedback loop includes the type of backlash described immediately above. Systems 1 and 2 are shown in Fig. 4. The details of the hysteresis blocks are as shown in Fig. 2.

The input signal to System 1 is white noise, filtered as shown, to derive $r(t)$, an element of a random, stationary-ergodic, Gaussian process. In System 1, the half dead-space size is $b=1.30$ volts, $[\phi_{rr}(0)]^{1/2}$ is 10.0 volts rms, $T_2=1.0$, and $T_1=1.50$. Since

$$\phi_{rr}(0)=\frac{N^2}{3}$$

then

$$N^2=300$$

From well-known linear analysis procedures, the mean-square value of the signal $y(t)$ (the input to the dead-zone block) is given by

$$\phi_{yy}(0)=\frac{1.25N^2}{7.125(K_{dz}C_H)^2+2.50K_{dz}C_H} \tag{10}$$

The selection of C_H involved a short study in itself. For this purpose, System 1 was simulated on the Wisconsin-Philbrick analog computer described in reference 1 (using a Philbrick dead-zone unit) and oscilloscope plots of $i(t)$ versus $f(t)$ were studied. A value of 15 for C_H gave a hysteresis characteristic which appeared satisfactory at the frequencies involved.

Putting $d=1.0$, $b=1.30$, and $C_H=15.0$, K_{dz}/d versus $[\phi_{yy}(0)]^{1/2}b=\beta/b$ is plotted in Fig. 3 (the "system" plot), and it is found that this intersects the curve specifying the gain of the dead-zone element at

$$K_{dz}=0.35 \tag{11}$$

hence for System 1

$$C_H K_{dz}=5.25 \tag{12}$$

The remaining part of the solution simple, linear analysis with the hyster block replaced by one enclosing a li device with frequency-response func specified by

$$\frac{F}{I}(j\omega)=\frac{C_H K_{dz}}{j\omega+C_H K_{dz}}$$

The result of this analysis of the qu linearized form of System 1 gives

$\tau<0$; $_n\phi_{rc}(\tau)=0.44\epsilon^{0.667\tau}$

$\tau>0$; $_n\phi_{rc}(\tau)=1.248\epsilon^{-0.667\tau}-0.009\epsilon^{-5.46}$
$0.792\epsilon^{-0.392\tau}[\cos 0.897\tau-1.085\sin 0.8$

where

$_n\phi_{rc}(\tau)$ indicates the quantity $\phi_{rc}(\tau)/\phi_r$

and

$_n\phi_{cc}(0)=0.911$

The input-output cross-correla function for System 1 was also de mined on the analog computer. equipment used here is essentially same as that for the measurements scribed in reference 1. In the compu the hysteresis characteristic is derive using the same C_H that was assumed the quasi-linearized analysis. Ana computer measurements are plotte Fig. 5, along with the curve for $_n\phi$ which is predicted by the quasi-li analysis. Comparison of the predi curve and the analog data illustrates

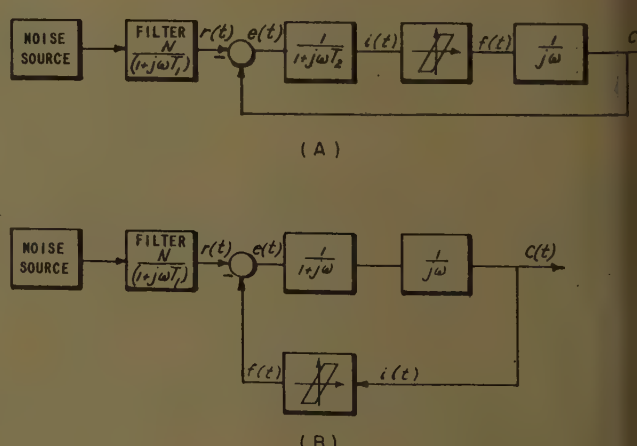
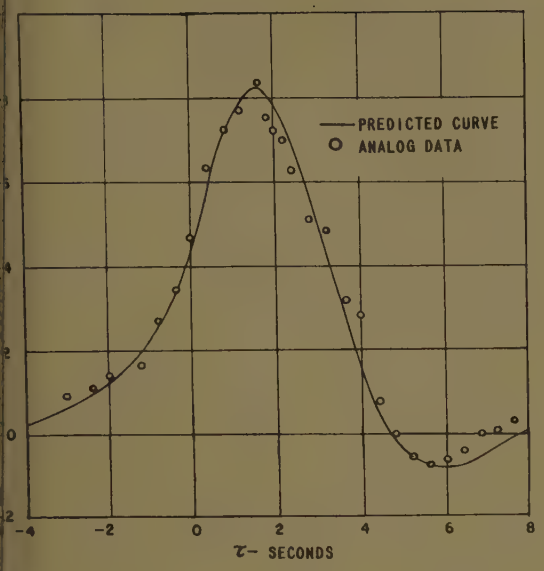


Fig. 4. Block diagrams for systems with hysteresis

A—System 1
B—System 2



g. 5. Input-output cross-correlation for System 1

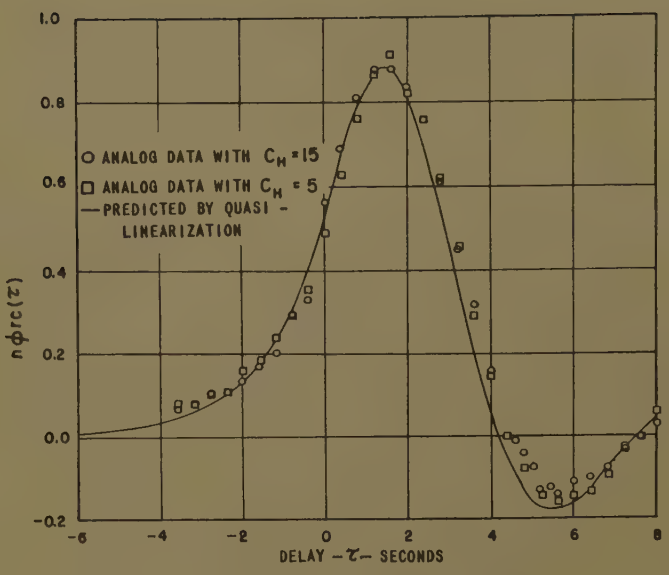


Fig. 6. Input-output cross-correlation for System 2

dity of the quasi-linear analysis of system with hysteresis. The meas- l value of $_n\phi_{cc}(0)$ is 0.94.

he analysis of System 2 is very lar to that for System 1. Here T_1 T_2 are as in System 1, but the size he dead zone is specified by $b=1.65$, the hysteresis element is included in feedback loop rather than as a part of ward loop. The rms value of the in- signal is, again,

$(0)]^{1/2} = 10.0$ volts

$=300$

analyzing the quasi-linearized system general K_{dz} and C_H it is found that

$$(0) = N^2 \times \frac{1.25+0.75K_{dz}C_H}{K_{dz}C_H(7.125K_{dz}^2C_H^2+2.50K_{dz}C_H)} \quad (15)$$

Analog-computer studies of the same e as were discussed for System 1 in- ated that a gain of $C_H=5.0$ was suffi- ntly high for System 2. With $C_H=$, the necessary system plots can be nstructed to show that

$_e=0.525$ (16)

e curve of $_n\phi_{rc}(\tau)$ for System 2 pre- ted by quasi-linearization is thus found analysis of the linear system obtained replacing the hysteresis block in Fig. 3) with an element having the fre- ency-response function

$$(j\omega) = \frac{2.62}{j\omega+2.62} \quad (17)$$

The input-output cross-correlation nction predicted by quasi-linearization System 2 is

$\tau<0$; $_n\phi_{rc}(\tau) = 0.523e^{0.667\tau}$

$\tau>0$; $_n\phi_{rc}(\tau) = 0.896e^{-0.667\tau} + 0.0076e^{-3.04\tau} - 0.379e^{-0.289\tau}[\cos 0.882\tau - 2.61 \sin 0.882\tau]$ (18)

In System 2, quasi-linearization gives

$_n\phi_{cc}(0) = 0.807$

The results of the analog-computer determination of the input-output cross-correlation for the actual nonlinear system are plotted along with the predicted curve in Fig. 6. In the analog-computer study, the hysteresis characteristic is again derived by using a dead-zone element and high-gain integrator loop, but in the computer a higher value of C_H ($C_H=15$) was used. In Fig. 6, the smooth curve is the predicted $_n\phi_{rc}(\tau)$ for System 2, and the data points enclosed in circles are analog-computer data for System 2 with $C_H=15$. To check that the value for C_H of 5.00 used in the quasi-linear analysis is adequate, analog-computer measurements of $_n\phi_{rc}(\tau)$ were included for this case. If the value of 5.00 is adequate, there should be no appreciable difference between the two sets of analog-computer data. As is seen by reference to Fig. 6, there is only a small difference between the two sets of experimental data and both match the predicted curve quite well. When measured, $_n\phi_{cc}(0)$ was found to be 0.83.

Servomechanisms Incorporating Relays With Hysteresis

The methods of the preceding section will now be applied to determine the input-output cross-correlation function for a relay servomechanism in which the relay has the characteristic shown in Fig. 7. This characteristic may be developed by cascading an idealized-relay element with a hysteresis element having the characteristic shown in Fig. 1. The proper arrangement of these elements is shown in Fig. 8 with the hysteresis element again represented by the closed-loop configuration used earlier. The relay block representation now contains two zero-memory nonlinear elements, in the quasi-linear approximation, there will be two gain constants K_{dz} and K_{re} which are initially undetermined.

The general method of solution is graphical and straightforward. In operational notation, the frequency-response function for the linearized relay block is given by

$$\frac{C'}{R'}(j\omega) = \frac{K_{re}K_{dz}C_H}{j\omega+K_{dz}C_H} \quad (19)$$

For clarity, the general method of solution will be outlined, after which a specific example is considered. The step-by-step procedure for determination of the input-output cross-correlation function for a system of this type is as follows:

1. Plot the curve showing equivalent-gain variation of the idealized relay as a function of rms input to the idealized relay (σ) (see reference 1).

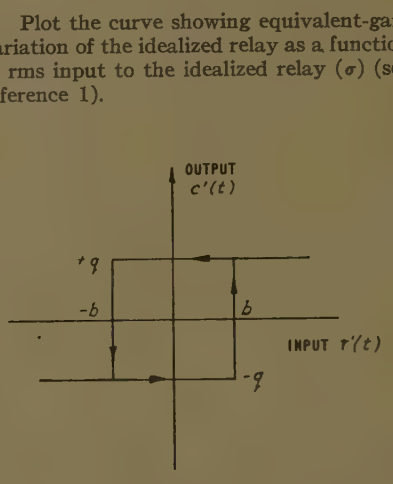


Fig. 7. Relay hysteresis characteristic

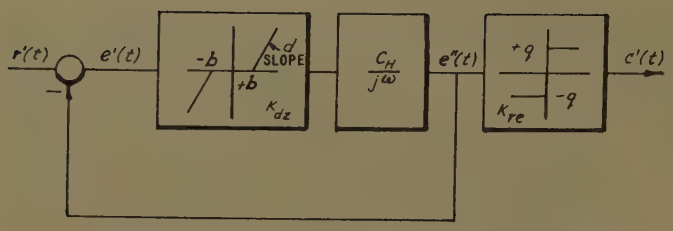


Fig. 8. Computer simulation of relay with hysteresis

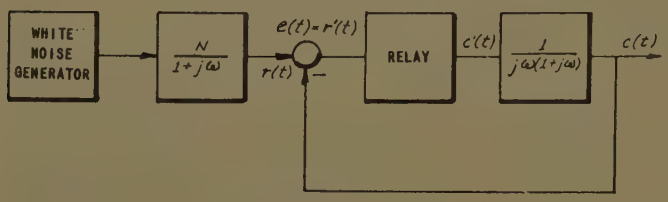


Fig. 9. Relay servo, System 3

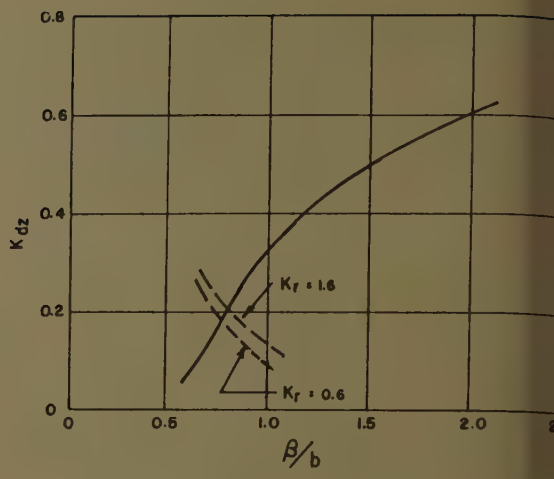


Fig. 10; Gain of deadzone element and system pl for System 3

2. Plot, on axes separate from the plot of step 1, the curve showing variation of K_{dz}/d as a function of the ratio of rms input to the dead-zone element, to the half-width of the hysteresis loop (β/b); i.e., reproduce the solid curve of Fig. 3.

3. Choose C_H.

4. Solve the quasi-linearized system obtained by inserting the detailed relay-block into the system, finding:

(a) $\phi_{e'e'}(0)$ in terms of the system parameters, K_{dz} and K_{re}.

(b) $\phi_{e''e''}(0)$ in terms of the system parameters, K_{dz} and K_{re}.

Note that

$$\phi_{e'e'}(0)=\beta^2$$

and

$$\phi_{e''e''}(0)=\sigma^2$$

5. For a set of values of K_{re} selected at intervals over the range of variation of K_{re}, find system plots with respect to K_{dz} and add these to the gain curve of step 2. The intersections of these system plots with the curve plotted in step 2 specify the values of K_{dz} which must accompany each one of the set of K_{re}'s.

6. Plot the intersection points of step 5 on a set of axes with K_{dz} as abcissas and K_{re} as ordinates. Draw a smooth curve through the plotted points.

7. Select an arbitrary set of K_{re}'s. For each find the required value of K_{dz} by reference to the curve of step 6. From the expression for $\phi_{e''e''}(0)=\sigma^2$ of step 4, construct a system plot with respect to K_{re} and add this curve to the gain curve of step 1. The intersection of these curves gives a value of K_{re}, and by reference to the plot of step 6, a value of K_{dz}, which are such as to satisfy simultaneously the compatibility requirements for K_{dz}, K_{re}, $\phi_{e'e'}(0)$ and $\phi_{e''e''}(0)$.

8. Using the prescribed values for K_{dz} and K_{re}, solve the quasi-linearized system by linear analysis for $\phi_{rc}(\tau)$ and $\phi_{cc}(0)$.

As an example illustrating these steps, $_n\phi_{rc}(\tau)$ will be determined for the system shown in Fig. 9 after which this solution will be checked by use of the analog computer. System 3 has the input signal spectral density

$$\Phi_{rr}(\omega)=\frac{N^2}{1+\omega^2} \tag{20}$$

and input-signal mean-square value

$$\phi_{rr}(0)=\frac{N^2}{2} \tag{21}$$

If the input rms level is

$$[\phi_{rr}(0)]^{1/2}=10 \text{ volts}$$

then

$$N^2=200$$

For this level of input signal a reasonable relay-characteristic is specified by $/q/=10.0$ volts; $b=6.0$ volts.

The gain plots of steps 1 and 2 are shown as the solid curves of Figs. 10 and 11. System 3 has about the same time constants as the systems with hysteresis studied in earlier sections, but it has only a single "break" in the attenuation plot for the input shaping-filter, so that $r(t)$ has more high-frequency content than does System 1, for example. It will be assumed that $C_H=25.0$ is a suitable with a later check for confirmation.

Inserting the detailed block of Fig. 8 into the relay block of Fig. 9 and solving for $\phi_{e'e'}(0)$ and $\phi_{e''e''}(0)$ (step 4), gives

$$G_1(j\omega)=\frac{E^1}{R}(j\omega)$$

$$=\frac{(j\omega)^2(1+j\omega)}{j\omega(1+j\omega)(j\omega+C_HK_{dz})+C_HK_{dz}K_{re}} \tag{22}$$

$$\Phi_{e'e'}(\omega)=\Phi_{rr}(\omega)/G_1(j\omega)/^2 \tag{23}$$

$$\Phi_{e'e'}(\omega)\times$$

$$\left|\frac{N(jw)^2}{j\omega(1+j\omega)(j\omega+C_HK_{dz})+C_KK_{dz}K_{re}}\right|^2 \tag{24}$$

$$\phi_{e'e'}(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_{e'e'}(\omega)d\omega$$

$$=\frac{N^2}{1+C_HK_{dz}-K_{re}}$$

Similarly,

$$\phi_{e''e''}(0)=\frac{N^2}{2}\left[\frac{C_HK_{dz}}{(C_HK_{dz}+1)-K_{re}}\right] \tag{2}$$

With equation 24, and using $C_H=25$. system plots with respect to K_{dz} are co structed for $K_{re}=0.6$, 0.8, 1.0, 1.2, 1. and 1.6. Two of these system plots ha been added to Fig. 10 (the dotted curves Noting that K_{dz} is between 0.18 and 0. for any value of K_{re} in the range from 0 to 1.6, and that the curve for K_{dz} nearly linear over this small range, t plot of step 6 is omitted, and it is a sumed that linear interpolation will satisfactory.

The system plot with respect to K_{re} shown as the dotted curve in Fig. The point of intersection gives the follow ing information for System 3.

$K_{re}=0.8$

$\sigma/q=0.98$

$K_{dz}=0.19$

$C_HK_{dz}=4.75$

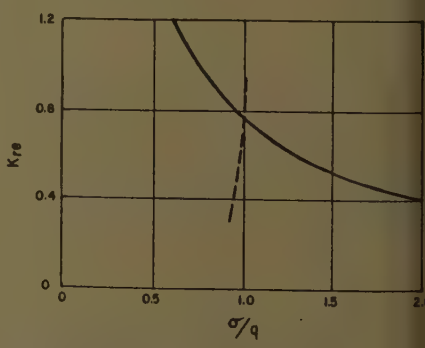


Fig. 11. Gain of relay and system plot f System 3

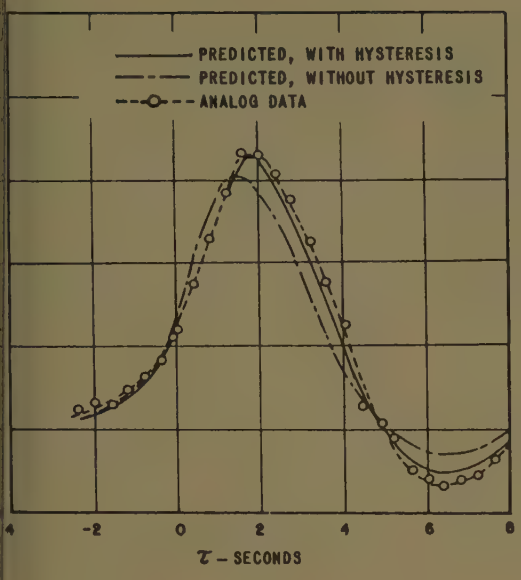


Fig. 12. Input-output cross-correlations for System 3

h these values, the quasi-linearized of System 3 is easily solved by analysis to find that the normalized output cross-correlation function is by:

$$_n\phi_{rc}(\tau) = 0.251e^{\tau}$$

$$_n\phi_{rc}(\tau) = 1.016e^{-\tau} - 0.0153e^{-4.947\tau} - 747e^{-0.405\tau}[\cos 0.776\tau - 1.55 \sin 0.776\tau] \quad (27)$$

$_n\phi_{cc}(0)$.

e analog-computer simulation for em 3 involves special problems. l amounts of zero drift or offset in ignal $e''(t)$ radically affect the cascade arp limiters and high-gain amplifiers to provide the idealized-relay char-istics. It is necessary, therefore, ercise extreme care in adjustment of the nonlinear Philbrick units, which are not chopper-stabilized. In spite of these difficulties, the curve of $_n\phi_{rc}(\tau)$ was determined for System 3. Fig. 12 includes three curves. The solid-line curve is the input-output cross-correlation function predicted by quasi-linearization theory. The dotted curve represents a smoothing of the data measured on the analog computer for the nonlinear system. For comparison, a curve of $_n\phi_{rc}(\tau)$ (computed on the basis of quasi-linearization) when no relay hysteresis is considered is included. Note that the method of this paper correctly predicts the amount of shift in the first peak of $_n\phi_{rc}(\tau)$ and the amount of increase in the first peak of this curve; both of these effects are the result of the relay hysteresis. The measured value for $_n\phi_{cc}(0)$ is 0.60. During the analog computer solution of this problem observation of the input-output characteristic of the hysteresis element confirmed that the value chosen for C_H was satisfactory.

Conclusions

The method of quasi-linearization has been found correctly to predict the effects of hysteresis-type nonlinearities on the input-output cross-correlation functions of several different closed-loop control systems, under the limitation of Gaussian input signals. In the quasi-linear models of the nonlinear systems, the hysteresis element is represented by a first-order lag element.

The inclusion of relay hysteresis in a second-order relay servo was found to result in the first maximum of the input-output cross-correlation curve being lowered and shifted in the direction of increased output delay. The necessary curves, and a general method, for studying the effects of relay hysteresis were presented.

References

1. Generalized Error Functions for Zero-Memory Nonlinear Systems with Random Inputs, H. R. Leland. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 14, 1958.

2. Nonlinear Control Systems with Random Inputs, R. C. Booton, Jr. *Report no. 61*, Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Mar. 1952.

3. Tables of Functions (book), E. Jahnke, F. Emde. Dover Publications, Inc., New York, N. Y., 1945.

4. Role of Diodes in an Electronic Differential Analyzer. *Report GER-4798*, Goodyear Aircraft Corporation, Akron, Ohio, April 1952.

Optimization of the Adaptive Function by Z-Transform Method

S. S. L. CHANG
MEMBER AIEE

psis: A study is made on the optimiza- of the adjustment process in a self- nizing system under rather general nptions. The system is designed to a performance parameter m either at rescribed value or at an unknown mal value. A direct measurement on made and the adjustment is based on sured m only.

ctors considered are: 1. The finite suring interval; 2. The necessity of ing ahead one interval; 3. The able error in measurement; and 4. changing situation. A set of weighting rs on present and past data, and the er value of test bias for extremal-seeking systems, are determined by a least square optimization process. The criterion of optimization is least reduction in m for peak-seeking systems and least square error in m for systems with prescribed value of m.

Two types of extremal-seeking systems are studied. The alternate biasing systems are found to be superior in performance compared to the derivative sensing systems.

A SYSTEM may be characterized as self-optimizing under the following conditions:

1. The system operates under changing situation.

2. The system has a set of adjustable parameters.

3. The ultimate purpose is to keep a set of system performance parameters at specified or extremal values by automatically adjusting the set of adjustable parameters.

The foregoing conditions can be considered sufficient for a control system to

Paper 59-1296, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE-ISA National Automatic Control Conference, Dallas, Texas, November 4–6, 1959, and re-presented at the AIEE Winter General Meeting, New York, N. Y., January 31–February 1, 1960. Manuscript submitted May 26, 1959; made available for printing November 27, 1959.

S. S. L. Chang is with New York University, New York, N. Y.

The research reported in this paper was sponsored by the U. S. Air Force Office of Scientific Research, Air Research and Development Command, under Contract No. *AF 49(638)-586*.

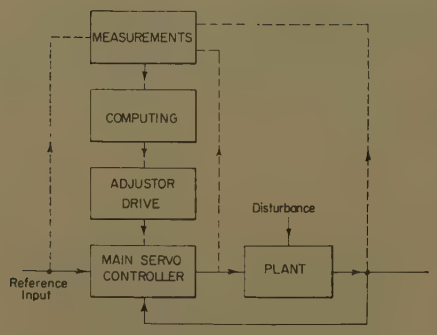


Fig. 1. Block diagram of a self-optimizing system

be called self-optimizing in the sense that, once these conditions are met, it is generally agreed that it is a self-optimizing system.

To be more specific, the changing situation may represent a change in the statistical properties of the inputs, either desired or undesired, or a change in the transfer function of controlled plant. The performance parameters to be kept at specified values may be system bandwidth, damping factor, or mean square value of control effort. The performance parameters to be kept at extremal values may be engine efficiency, per-cent yield in a chemical process, and the mean square value of system error, etc.

Fig. 1 illustrates the two types of signal flow in such a control system. The main servo controller receives input signal and exercises direct control over the plant operation. For any particular setting of the adjustable parameters in the main servo controller, the closed-loop system as connected by the solid lines operates in exactly the same way as an ordinary feedback control system. The broken lines represent signal flow only for the purpose of making adjustment on the adjustable parameters. This measuring and adjusting operation is a control system in itself and will be referred to as adaptive control.

The adjustments may be based on two types of data:

1. Measurements on changing situation.
2. Measurements on performance parameters.

If the data of type 1 are used, the adaptive control is an open-loop system. If the data of type 2 are used, the adaptive control is a closed-loop system as the adjustments change the performance parameters on which the adjustments are based. Consequently the latter type has all the inherent advantages of a closed-loop system: It is self-correcting and can be instrumented with relatively less hardware requirements.

In the literature most papers are concerned with the determination, or estimation, of situation and performance parameters and/or showing that a particular adaptive loop is stable or tolerably oscillatory.[1-11] Admittedly, under operating conditions with little or no change in the situation parameters, a stable adaptive control loop or one with a tolerable amount of hunting is all that is necessary. However, there are applications with rapid variations in the situation parameters and the ability of the system to track such variations becomes a vital consideration. Some of the general problems are:

1. Measurement. As stochastic signal of one form or another is present, the probable error in measurement is large if a very short interval is used. On the other hand, if a long interval is used, the measured result, which is a product of past situation, would not adequately reflect the present situation. A compromise in the promptness and accuracy has to be made. What represents the best compromise?

2. Adjustment. The adjustment is on future operation while it is based on present and past data only. Furthermore, the data are inaccurate because of the compromise mentioned in problem 1. Considering all these factors, the current data are not necessarily the sole criteria for adjustment in an optimum system. What weights should be assigned to previous data in computing the required adjustments?

3. Evaluation. Given a statistical description of the changing situation, and probable error in measurement, how close can one hold the performance parameters to specified values or to their respective extremal values?

Preliminary to a solution of the foregoing problems is a criterion for the best adaptive control system. For extremal-seeking systems, a natural criterion is the difference between the extremal value and the average value of the performance parameter. For systems which regulate a performance parameter to a specified value, the mean square value of the deviation is used.

In this report, a solution of problems 2 and 3 are given for systems which base their adjustment operations on a directly measured performance parameter. The situation parameters are not known or measured. The approach is as follows: Given the values of the interval and probable error of measurement, and some statistical description of the changing situation, the optimum weighting factors on present and past data are determined. The deviation from the extremal valu the mean square deviation from the gi value is then calculated. It is the lo obtainable for a given way of meas ment. To determine the best metho measurement (problem 1) and the o mum T to use, one has to know the depe ence of probable measuring error on which differs from one problem to anot However, knowing this dependence the solution to problem 3, the solutio problem 1 becomes a matter of calculu

In a given system, there are a num of adjustable parameters among wh a lesser number are independent. bring forth the underlying conce with the least amount of mathemat only systems with one independ adjustable parameter are studied in present report.

Nomenclature

a = a constant of the adaptive control lo
$b = (\partial m/\partial x)x\alpha$ for a system with speci parameter m
$b' = -(\partial^2 m/\partial x^2)x\alpha$ for a peak seeking sys
C_1 = mean square value of x_α
$C_1' = 2\nu T C_1$
D = deviation function. $D = \overline{(m_s - m_i)^2}$ a system with specified m_i, $D = \overline{m_{opt} - m_i}$ for a peak seek system
e = error in adjustable parameter. $e = x_\alpha$
$f_D(\eta)$, $f_q(\eta)$ = dimensionless functions calculating D and q^2 in terms of
$F(z)$ = unknown system function of adap loop, to be determined
i, j, l, n = running indexes
$K(z)$ = closed loop system function of adaptive loop
m_i = performance parameter at ith samp instant
m_i' = measured value of m_i
m_s = specified value of m_i same for all
m_{opt} = optimum value of m_i. The valu m_{opt} depends on i
q = test bias for peak-seeking systems
s = Laplace transform variable
t = time
T = sampling interval. T is also measuring interval except for der tive sensing systems in which the measuring interval is denote T_1
u_i = measured difference $m_s - m_i'$ in syst with specified m, $m_{i1}' - m_{i2}'$ derivative sensing systems, $(-1)^i m_i'$ in alternate biasing syst
w_i = weighting factor of the nth previous
x = adjustable parameter
x_i = the value of x at ith sampling inter
x_i' = actually applied value of x_i in mea ing intervals of peak seeking syst $x_i' = x_i + q$
x_α = ideal value of x to give $D = 0$. averaged value of x_α over a period of time is zero. If it is $\bar{x}_\alpha$ should be subtracted from all x's so that this condition is not
$Y(z)$ = a rational function of z with its poles and zeroes inside the circle
z = the z-transform variable e^{sT}

ation parameter
suring error. $m_i' = m_i + \delta_i$
-transform of δ
n square value of δ
), $\Phi_{xx}(s)$, $\Phi_{\delta\delta}(s)$ = power spectra or spectral density functions of the respective variables
), $\Phi_{xx}(z)$, $\Phi_{\delta\delta}(z)$ = sampled power spectra
constant of the auto-correlation function of x_α
function of z which is analytic in the unit circle
imensionless parameter proportional to Δ

ms with Specified m

Assumptions

α_i, m_i, and x_i denote the situation eter, performance parameter, and justable parameter respectively at n measuring interval. The meas- erformance parameter m_i' can be n as

$$m_i(\alpha_1, \alpha_2 \ldots \alpha_{i-1}, \alpha_i; x_1, x_2 \ldots x_{i-1}, x_i) + \delta_i \quad (1)$$

value of m_i depends on the present l as previous values of α and x, and random variable representing error easurement. Let m_s denote the ed value of m_i. A condition repre- g adequate coverage by x of all le situations α can be stated as s:

every sequence $\alpha_1, \alpha_2 \ldots \alpha_i$ with ro probability, there exists a se- e $x_{\alpha 1}, x_{\alpha 2} \ldots x_{\alpha i}$ such that

$$\alpha_2 \ldots \alpha_{j-1}, \alpha_j; x_{\alpha 1}, x_{\alpha 2} \ldots x_{\alpha(j-1)}, x_{\alpha j}) = m_s \quad (2)$$

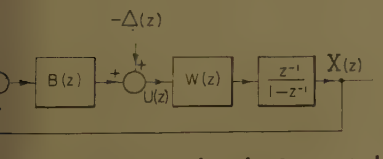
$j = 1, 2 \ldots, i$.

s condition is necessary if it is at ssible to hold m_j to the specified m_s, and will be assumed to be

will be assumed that the tracking ss is fairly good so that $x_{\alpha i} - x_i$ is enough for the following approxima- o be good.

$$\alpha_2, \ldots \alpha_{i-1}, \alpha_i; x_1, x_2 \ldots x_i)$$
$$= m_i(\alpha_1, \alpha_2 \ldots \alpha_i; x_{\alpha 1}, x_{\alpha 2} \ldots x_{\alpha i}) + \sum_{j=1}^{i} \left(\frac{\partial m_i}{\partial x_j}\right)_\alpha (x_j - x_{\alpha j}) \quad (3)$$

the partial derivatives $\partial m_i / \partial x_j$ are ated at $x_j = x_{\alpha j}$, $j = 1, 2, \ldots i$. ese derivatives depend on the α's



Block diagram of the adaptive control loop

only. Usually the time interval of measurement is greater than the response time of the main servo system, and m_i depends mostly on x_i, slightly on x_{i-1}, and not at all on the previous x's.

As the situation parameter α is unknown, the exact values of the partial derivatives $\partial m_i / \partial x_j$ are not known. The proposed procedure is based on a set of nominal values b_n such that

$$\left(\frac{\partial m_i}{\partial x_j}\right)_\alpha = b_{i-j} \quad (4)$$

This assumption can be justified later by examining the resulting adaptive control loop.

The adjustment procedure is represented by the following equation:

$$x_{i+1} = x_i + \sum_{j=1}^{i} w_{i-j}(m_s - m_j') \quad (5)$$

where w_n, $n = 0, 1, 2 \ldots$ are the weighting factors attached to present and past measurements. The optimum values of w_n are to be determined.

The preceding assumptions can be summarized as:

1. Condition of adequate coverage: equation 2.
2. Linearization: From equations 1 to 4,

$$m_i' = m_s + \sum_{j=1}^{i} b_{i-j}(x_j - x_{\alpha j}) + \delta_i \quad (6)$$

3. Adjustment procedure

$$x_{i+1} = x_i + \sum_{i=1}^{j} w_{i-j}(m_s - m_j') \quad (7)$$

Z-Transform Representation

Let the measured difference u_i be defined as

$$u_i = m_s - m_i' \quad (8)$$

If equation 6 is multipled by z^{-i} and sum over i, it becomes

$$U(z) + \Delta(z) = \sum_{i=0}^{\infty} \sum_{j=1}^{i} b_{i-j}(x_{\alpha j} - x_j) z^{-i}$$
$$= \sum_{j=0}^{\infty} \sum_{i-j=0}^{\infty} b_{i-j} z^{-(i-j)} (x_\alpha - x_j) z^{-j}$$
$$= B(z)[X_\alpha(z) - X(z)] \quad (9)$$

where $U(z)$, $\Delta(z)$, $B(z)$, $X_\alpha(z)$, $X(z)$ are the corresponding z-transforms.

If equation 7 is multiplied by z^{-i-1} and sum over i, it becomes

$$X(z) = z^{-1} X(z) + z^{-1} W(z) U(z) \quad (10)$$

where $W(z)$ is the z-transform of the weighting factors. Equation 10 can be rewritten as

$$X(z) = \frac{z^{-1}}{1 - z^{-1}} W(z) U(z) \quad (11)$$

Equations 9 and 10 are represented by the block diagram of Fig. 2.

Optimization of Adaptive Control Loop

In Fig. 2, the only unknown function is $W(z)$. This function is to be selected such that

$$D \equiv \overline{(m_s - m_1)^2} = \text{minimum} \quad (12)$$

Let $\bar{u}_i$ represent $m_s - m_i$. Equation 9 gives

$$\bar{U}(z) = B(z)[X_\alpha(z) - X(z)] \quad (13)$$

Equations 12 and 13, and Fig. 2 specify the optimization problems completely. Using the method of reference 12 one obtains:

$$W(z) = \frac{(z-1)K(z)}{B(z)[1 - K(z)]} \quad (14)$$

where $K(z)$ is the closed-loop system function of the adaptive control loop. Its optimum form is

$$K(z) = \frac{1}{Y(z)} \left[\frac{B(z) B(z^{-1}) \Phi_{z_\alpha z_\alpha}(z)}{Y(z^{-1})} \right]_i \quad (15)$$

In equation 15, $Y(z)$ is defined by the following two conditions:

1. $Y(z) Y(z^{-1}) = B(z) B(z^{-1}) \Phi_{x_\alpha x_\alpha}(z) + \Phi_{\delta\delta}(z)$ (16)
2. All the poles and zeroes of $Y(z)$ are inside the unit circle.

The symbol $[\ ]_i$ is defined as follows: Any rational function $Q(z)$ can be expressed as

$$Q(z) = \sum_j \frac{A_j}{z - a_j} + \sum_j \frac{B_j}{z - b_j} + \text{polynomial in } z$$

where $a_j < 1$ and $b_j > 1$. Then

$$[Q(z)]_i \sum_j \frac{A_j}{z - a_j} \quad (17)$$

From $K(z)$, D can be calculated as

$$D(z) = \frac{1}{2\pi j} \oint \{B(z) B(z^{-1}) \Phi_{x_\alpha x_\alpha}(z) \times [1 - K(z)][1 - K(z^{-1})] + K(z) K(z^{-1}) \Phi_{\delta\delta}(z)\} \frac{dz}{z} \quad (18)$$

Alternative Optimization Procedure

Sometimes the change in α is in distinctive steps, or ramps, and the condition for designing the adaptive control loop is:

1. Under steady-state condition the mean square error is limited to a low value:

$\overline{(m_s-m_t)^2}$ steady state

$$=\frac{1}{2\pi j}\oint K(z)K(z^{-1})\Phi_{\delta\delta}(z)\frac{dz}{z}\leq L \tag{19}$$

2. The integral square error during the transient interval is a minimum

$$\overline{(m_s-m_t)^2}_{\text{transient}}=\frac{1}{2\pi j}\oint B(z)B(z^{-1})\times X_\alpha(z)X_\alpha(z^{-1})[1-K(z)][1-K(z^{-1})]\frac{dz}{z}$$
$$=\text{minimum} \tag{20}$$

Using Lagrange's undetermined multiplier k^2, equations 19 and 20 are equivalent to

$$\overline{(m_s-m_t)^2}_{\text{transient}}+k^2\overline{(m_s-m_t)^2}_{s.s.}=\min \tag{21}$$

The solution to equation 21 is exactly the same as equation 15, except that now $X_\alpha(z)X_\alpha(z^{-1})$ takes the place of $\Phi_{x_\alpha x_\alpha}(z)$ and $k^2\Phi_{\delta\delta}(z)$ takes the place of $\Phi_{\delta\delta}(z)$.

The unknown constant k^2 is determined by

$$\frac{1}{2\pi j}\oint F(z)F(z^{-1})\Phi_{\delta\delta}(z)\frac{dz}{z}=L \tag{22}$$

Example 1. Sometimes the operating situation changes in sudden steps. Correspondingly x_α is also steplike, and its power spectrum can be approximated as

$$\Phi_{x_\alpha x_\alpha}(s)=\frac{2\nu C_1}{\nu^2-s^2} \tag{23}$$

$$\Phi_{x_\alpha x_\alpha}(z)=\frac{C(1-e^{-2\nu T})}{(1-e^{-\nu T}z^{-1})(1-e^{-\nu T}z)} \tag{24}$$

The random errors in successive measurements are assumed to be independent.

$$\Phi_{\delta\delta}(z)=\Delta \tag{25}$$

where Δ is the mean square value of δ_t. In case the response time of the main servo system is small compared to T, the constants b_n are approximately

$$b_0=b,\ b_i=0 \text{ for all } i\neq 0 \tag{26}$$

Equations 24 to 26 can be viewed as representing the starting point of a typical design problem. The foregoing arguments are simply aimed to show that this particular set of assumed values has some claim to generality.

The optimum weighting factors w_i and minimum value of D are to be determined.

Solution. By definition:

$$Y(z)\,Y(z^{-1})\equiv\frac{b^2C_1(1-e^{-2\nu T})}{(1-e^{-\nu T}z^{-1})(1-e^{-\nu T}z)}+\Delta$$
$$=\frac{M^2(1-az^{-1})(1-az)}{(1-e^{-\nu T}z^{-1})(1-e^{-\nu T}z)} \tag{27}$$

where the constant a is given by

$$a+\frac{1}{a}=2\cosh\nu T+\frac{2b^2C_1}{\Delta}\sinh\nu T \tag{28}$$

$a<1$

and the constant M is given as

$$M^2=\frac{b^2C_1(1-e^{-2\nu T})}{(1-ae^{-\nu T})(1-ae^{\nu T})} \tag{29}$$

It follows from equation 27 and condition 2 that

$$Y(z)=\frac{M(1-az^{-1})}{1-e^{\nu T}z^{-1}} \tag{30}$$

From equations 17, 24, 29, and 30, one obtains

$$\left[\frac{B(z)B(z^{-1})\Phi_{x_\alpha x_\alpha}(z)}{Y(z^{-1})}\right]_t=\frac{M(e^{-\delta T}-a)}{z-e^{-\nu T}} \tag{31}$$

Equations 15, 30, and 31 give

$$K(z)=\frac{(e^{-\nu T}-a)z^{-1}}{1-az^{-1}} \tag{32}$$

Knowing $K(z)$, $W(z)$ and D are calculated according to equations 14 and 18, the results are

$$W(z)=\frac{(e^{-\nu T}-a)(1-z^{-1})}{b(1-e^{-\nu T}z^{-1})} \tag{33}$$

$$D=\left(\frac{1}{a}e^{-\nu T}-1\right)\Delta \tag{34}$$

In case of perfect measurement, $\Delta=0$. However, as $1/a$ becomes infinity, the value of D is indeterminate. To evaluate this limit, let Δ be an arbitrarily small value ϵ, and equation 28 gives

$$\frac{1}{a}\simeq\frac{2b^2C_1}{\epsilon}\sinh\nu T \tag{35}$$

Substituting equation 35 into 34, one obtains

$$D_o\equiv\lim_{\Delta\to 0}D=b^2C_1(1-e^{-2\nu T}) \tag{36}$$

D_o represents the probable error involved in predicting the operating situation of the next period.

It has been pointed out[13] that inexact cancellation of poles or zeroes on or outside the unit circle can cause instability in the closed-loop system. As equation 33 requires cancellation of a pole at $z=1$, it obviously would present this difficulty. However this difficulty can be readily overcome by using $1-(1-\epsilon)z^{-1}$ instead of $1-z^{-1}$ in equation 33, where ϵ is a very small positive constant. As the pole at 1 arises out of the computation process, its value is exact. By drawing the pole towards a zero at $1-\epsilon$, the closed-loop pole is always inside the unit circle. In calculating the deviation function D and weighting factors w_i, ϵ is assumed to be negligible.

Next the weighting factors w_i are culated. Since

$$W(z)=\sum_{i=0}^{\infty}w_iz^{-i}$$

equation 33 gives:

$$w_0=\frac{e^{-\nu T}-a}{b}$$

$$w_n=-w_0(e^{\nu T}-1)e^{-n\nu T},\ n=1,2,3,\ldots$$

To facilitate further discussion, numerical examples are calculated. given values and calculated results listed in Table I.

Table I. Two Numerical Cases

		Case 1	Case
Given values.	νT	0.1	0.01
	Δ	0.2 b^2C_1	0.2 b^2
Calculated values.	a	0.380	0.789
	D	0.276 b^2C_1	0.071
	D_o	0.182 b^2C_1	0.019

The constant b^2C_1 is the mean sq variation in m without adaptive con It is used here to gage the mean sq error in measurement as well as the e tiveness of the adaptive control The deviation function D_o for the idea uation of perfect measurement is listed for comparison. In Case 1, variations in operating situation relatively fast and the measuring cannot be effectively reduced by a aging over a number of past sam meanwhile keeping up with the chan situation. In Case 2, Δ is 10 times la than D_o, and some averaging can be to reduce the random measuring erro

A surprising fact about equation that the weighting factors on past are negative. However a close exan tion of Fig. 2 will show that these we ing factors represent only the imme

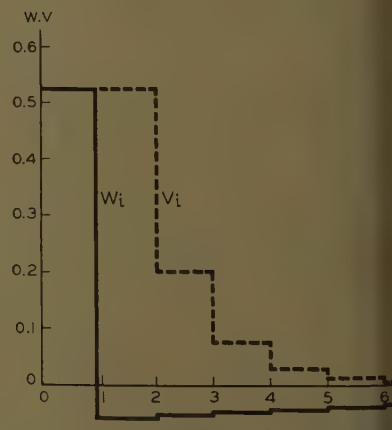
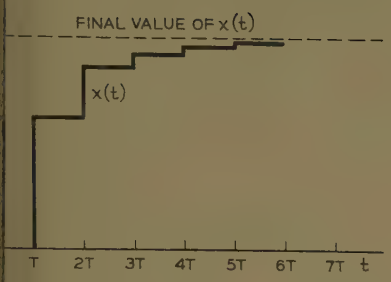


Fig. 3. The weighting factors w_i (solid and effective weighting factors v_i (b line) in units of 1/b



4. Transient response of the adaptive control loop to unit step at t=0

ibution. The total contribution of to the value of X(z) is given as

$$\equiv\sum_n v_n z^{-n}\equiv\frac{X(z)}{U(z)}=\frac{K(z)}{b}$$

$$\frac{e^{-\nu T}-a}{b}\quad\frac{z^{-1}}{1-az^{-1}}$$

efore

$$\frac{e^{-\nu T}-a}{b}\cdot a^{n-1},\ n=1,2,3\ldots \tag{38}$$

e factor V_n gives the total effect of a ured change of one unit at ith val to the value of x at $(i+n)$-th val, and will be called the effective nting factor. In Fig. 3, the imme- (solid line) and effective (broken weighting factors are plotted for 1.

nother interesting feature of the sys- can be illustrated by its transient nse to a step input in x_α, assum- hat $\Delta(z)=0$.

$$)=\frac{1}{1-z^{-1}}$$

$$=K(z)X_\alpha(z)=\frac{(e^{-\nu T}-a)z^{-1}}{(1-az^{-1})(1-z^{-1})}$$

$$e^{-\nu T}-a)(1+a+a^2+\ldots+a^{n-1})$$

$$=\frac{(e^{-\nu T}-a)(1-a^n)}{1-a} \tag{39}$$

e unit step response is plotted in Fig. t gradually approaches a final value 347 instead of unity. The reason is since the next value of x_α is un- lated to its present value, the im- ing change in x_α is more likely to be e opposite direction. The adaptive oller does not go all the way in ipation of this change.

OSTERIORI JUSTIFICATION OF THE NEARIZATION ASSUMPTION

the beginning of the analysis we have ned equation 6 together with a set of nominal values for b's, and a nominal value for $\Phi_{\delta\delta}(z)$. In most practical cases the b's and $\Phi_{\delta\delta}$ depend on α and the problem is not a linear one. However, it can be considered as a piecewise linear system, and the assumption of nominal values instead of using correct instantaneous values of the coefficients becomes a problem of mismatching. In this light, the effects of this assumption on various aspects of the adaptive control loop performance can be studied.

1. *Stability.* In a feedback control system, the forward components are made critical to system performance by the high orders of integration involved in the controlled plant. However this is not the case with the adaptive control loop of Fig. 2. The only integration involved is the $z^{-1}/(1-z^{-1})$ block. The presence of $B(z)$ does not appreciably affect this situation since the b_n's are a sequence of rapidly decreasing numbers. In the terminology of continuous systems, the time constants of $B(z)$ are small compared to T.

In the previous example, no matter what values of b and $\Phi_{\delta\delta}$ are assumed, $W(z)$ is of the form

$$W(z)=\frac{A(1-z^{-1})}{1-e^{-\nu T}z^{-1}}$$

The closed-loop pole is at $e^{-\nu T}-bA$. For the adaptive control to be unstable, we must have $bA>1+e^{-\nu T}$. A value of b 3.6 times its nominal value for Case 1, or 7.6 times its nominal value for Case 2, is necessary to cause instability. If the nominal value of b is selected somewhere near the high end in the range of values of b, instability never occurs.

2. *Increase in mean square error.* The increase in deviation function D due to temporary mismatching is generally not high for two reasons: (A). The closed-loop response is not as affected as the open-loop response; (B). At the vicinity of a minimum point, there is always a flat region. As an example, if the assumed nominal value of b or b_{nom} is twice the actual b, the resulting $W(z)$ is, instead of equation 33,

$$W(z)=\frac{(e^{-\nu T}-a)(1-z^{-1})}{2b(1-e^{-\nu T}z^{-1})} \tag{40}$$

In equation 40, b is the actual value. With the new value of $W(z)$, $K(z)$ and 1-$K(z)$ become

$$K(z)=\frac{e^{-\nu T}-a}{2}\quad\frac{z^{-1}}{1-\xi z^{-1}} \tag{41}$$

$$1-K(z)=\frac{1-e^{-\nu T}z^{-1}}{1-\xi z^{-1}} \tag{42}$$

where $\xi=(e^{-\nu T}+a)/2$. The function D is calculated as

$$\overline{(m_s-m_i)^2}=\frac{1}{2\pi j}\oint[K(z)K(z^{-1})\Phi_{\delta\delta}+[1-K(z)][1-K(z^{-1})]b^2\Phi_{z\alpha z\alpha}]\frac{dz}{z}$$

$$=\frac{1}{1-\xi^2}\left[\left(\frac{e^{-\nu T}-a}{2}\right)^2 f(T)+b^2C_1(1-e^{-2\nu T})\right] \tag{43}$$

Equation 43 gives 0.331 b^2C_1 for Case 1 and 0.089 b^2C_1 for Case 2.

For the foregoing examples, assuming that there is a 3 to 1 variation in b, a nominal value two thirds of the highest b may be used. Even at both extremes, the increase in D is no more than 10% that obtainable from an optimum system with varying weighting factors which are matched to every value of b.

3. *Economy.* A direct consequence of assuming constant values for the b's is a time invarient $W(z)$, which is simple to instrument.

REDUCTION TO CONTINUOUS CASE

Reduction to continuous case is possible, whenever the variations in situation parameter α are slow enough for the following two effects to be negligible:

A. The adjustment delay of one measuring interval.

B. The graininess of finite number of measured values being weighed or filtered in arriving at an incremental value of x. For instance, Case 1 is definitely discrete. A continuous approximation would be too erroneous to be useful. Case 2 is at the borderline region. An indication is that its first four or five effective weighting factors u_i have substantial values.

Assuming the effects A and B to be small, equations 6 and 7 become

$$m(t)=m_s+\int_o^t b(t-\tau)[x(\tau)-x_\alpha(\tau)]d\tau+\delta(t) \tag{44}$$

$$\frac{dx(t)}{dt}\cong\frac{x_{i+1}-x_i}{T}=\int_o^t w(t-\tau)[m_s-m(\tau)]d\tau \tag{45}$$

where in the limit of $T\to 0$, $n\to\infty$ and $nT=\tau$

$$b(\tau)=b(nT)=\frac{b_n}{T} \tag{46}$$

$$w(\tau)=w(nT)=\frac{w_n}{T^2} \tag{47}$$

If $u(\tau)\equiv m_s-m(\tau)$, equations 44 and 45 can be written as

$$U(s)=B(s)[X_\alpha(s)-X(s)]-\Delta(s) \tag{48}$$

$$X(s)=W(s)U(s) \tag{49}$$

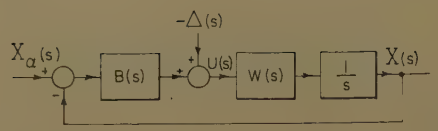


Fig. 5. Continuous approximation of a system with slowly varying situation

Equations 48 and 49 are represented by the block diagram of Fig. 5.

The spectral density $\Phi_{\delta\delta}(s)$ can be derived as follows: Assuming that each measurement is independent, the correlation function $\phi_{\delta\delta}(\tau)$ is a triangular pulse with height Δ and base $2T$. Therefore

$$\Phi_{\delta\delta}(s)=\int_{-\infty}^{\infty}\phi_{\delta\delta}(\tau)e^{-s\tau}d\tau=T\Delta \qquad (50)$$

Let $F(s)$ represent the closed loop system function

$$F(s)=\frac{\frac{1}{s}B(s)W(s)}{1+\frac{1}{s}B(s)W(s)} \qquad (51)$$

The deviation function D is

$$D=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\{B(s)B(-s)[1-F(s)]\times[1-F(-s)]\Phi_{x_\alpha x_\alpha}(s)+F(s)F(-s)\Phi_{\delta\delta}(s)\}ds \qquad (50)$$

Following well-known methods, the optimum form of $F(s)$ is obtained as

$$F(s)-\frac{1}{Y(s)}\left[\frac{B(s)B(-s)\Phi_{x_\alpha x_\alpha}(s)}{Y(-s)}\right]^{+} \qquad (51)$$

where $Y(s)$ is defined by two conditions

1. $Y(s)Y(-s)\equiv B(s)B(-s)\Phi_{x_\alpha x_\alpha}(s)+\Phi_{\delta\delta}(s)$ (52)

2. All the poles and zeroes of $Y(s)$ are in the left half plane. The symbol $[\ ]^{+}$ means partial fractioning the function inside and retaining only terms with poles in the left half plane.

Extremal Seeking Systems

Basic Assumptions

The parameters α_i, m_i, x_i, m_i', δ_i, are defined as before. The performance parameter m_i is assumed to depend on α_i and x_i only, and

$$m_i'=m_i(\alpha_i, x_i)+\delta_i \qquad (53)$$

In contrast to the previous case, m_i is to be held at either an unknown maximum or minimum, depending on the problem. With no loss of generality, it is assumed here that m_i is held at maximum.

The condition representing adequate coverage is stated as follows:

For every d_i, there exists a $x_{\alpha i}$ such that

$$m_i(\alpha_i, x_i)=\text{maximum}=m_{opt} \qquad (54)$$

The optimum value m_{opt} is not known. Equation 54 defines $x_{\alpha i}$. It is further assumed that the tracking process is fairly good such that $x_{\alpha i}-x_i$ is small enough for the following approximation to be good.

$$m_i(\alpha_i, x_i)=m_i(\alpha_i, x_{\alpha i})+\frac{1}{2}\left(\frac{\partial^2 m_s}{\partial x_i^2}\right)_{x_{\alpha i}}(x_{\alpha i}-x_i)^2 \qquad (55)$$

A nominal value $-b'$ is selected for the second derivative $(\partial^2\bar{m}_i/\partial x_i^2)x_{\alpha i}$, and equation 55 becomes

$$m_i(\alpha_i, x_i)=m_{opt}-\frac{b'}{2}(x_{\alpha i}-x_i)^2 \qquad (56)$$

The desired result is a procedure for adjusting x_i such that

$$D=\overline{m_{opt}-m_i}=\frac{b'}{2}\overline{(x_{\alpha i}-x_i)^2}=\text{minimum} \qquad (57)$$

Hunting Procedure

One characteristic aspect of extremal sensing systems is the hunting procedure. Even with negligible measuring error δ, it is not possible to tell whether m_i is at its peak without introducing some intentional variation in x. The manner in which this variation is made is called the hunting procedure.

Draper and Li[1] have suggested a number of hunting procedures in their work. However, one of their basic assumptions appears to be that the performance parameter can be measured instantaneously and that this measured value depends only on the present value of x. In a relatively fast-varying system, if the measuring interval is made long enough to insure that the measured m is sufficiently independent of α and x outside this interval, the measurements can no longer be considered continuous since α is likely to change substantially within a few such intervals.

Two hunting procedures which recognize the discrete nature of the problem will be analyzed:

1. *Derivative Sensing.* The adjustable parameter x of a typical derivative sensing system is illustrated in Fig. 6(A). At the end of the ith interval, x is adjusted to x_i+q and x_i-q in succession for smaller intervals T_1, and the subsequent x_{i+1} is determined by $m_{i1}'-m_{i2}'$ measured at the two intervals T_1 as well as all such previous measurements. This method is currently in use in chemical industry to optimize yield with satisfactory results. However, for systems with relatively fast changing situation, the following proposed method would be preferable.

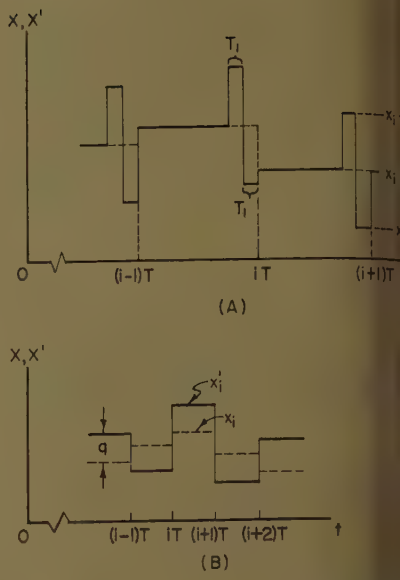


Fig. 6. Best estimate of adjustable param setting x_i (broken line) and actual adjust parameter setting x_i (solid line) of the types of extremal seeking systems

A—Derivative sensing
B—Alternate biasing

2. *Alternate Biasing.* The best e mate of the adjustable parameter x illustrated as the broken lines in 6(B). However instead of using estimated value x_i, the actual x_i' is justed to $x_i \pm q$, which is illustrated as solid lines. One obvious advantage this method is that measured data every interval is utilized on equal foot

Analysis of Derivative Sensing Systems

The changes in x_α and m_{opt} between two measuring intervals T are assu negligible. From equations 53 and 5

$$m_{i1}'=m_{opt}-\frac{b'}{2}(x_{\alpha i}-x_i-q)^2+\delta_{i1}$$

$$m_{i1}'=m_{opt}-\frac{b'}{2}(x_{\alpha i}-x_i+q)^2+\delta_{i2}$$

The difference of the two equatio

$$m_{i1}'-m_{i2}'=2b'q(x_{\alpha i}-x_i)+\delta_{i1}-\delta_{i2}$$

The correction in x_i is given as weighted sum of all previously meas differences:

$$x_{i+1}=x_i+\sum_{j=0}^{j=i}w_{i-j}(m_{j1}'-m_{j2}')$$

Equations 58 and 59 are identical equations 6 and 7 if

$$u_i=m_{i1}'-m_{i2}'$$

$$\delta_i=\delta_{i2}-\delta_{i1}$$

$$b_0=2b'q,\ b_s=0, \text{ for all } i\neq 0$$

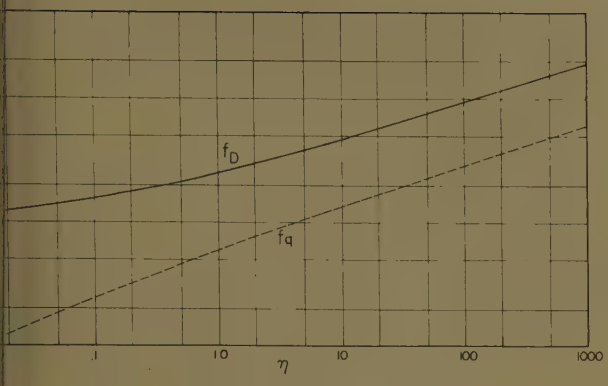


Fig. 7. The functions f_D and f_q versus η

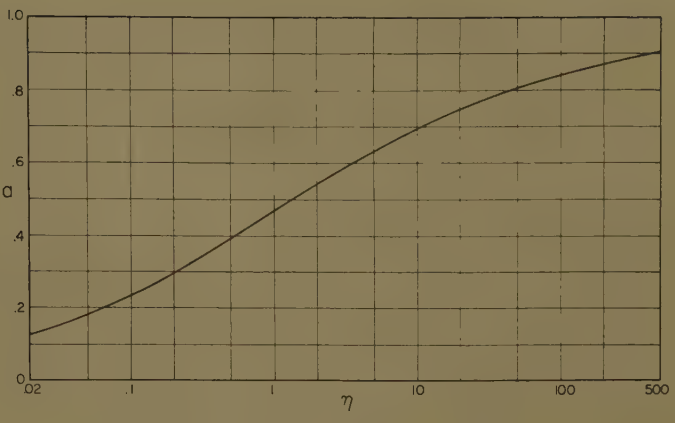


Fig. 8. The parameter a versus η

uently all the analytical results of ceding section hold true.

average reduction in m is

$$(T-2T_1)\overline{(x_{\alpha i}-x_i)^2}+$$

$$\frac{b'T_1}{2T}\overline{[(x_{\alpha i}-x_i-q)^2+(x_{\alpha i}-x_i+q)^2]}$$

$$\overline{(x_{\alpha i}-x_i)^2}+\frac{b'T_1q^2}{T} \tag{60}$$

n optimum $K(z)$, the value of $\overline{x_i)^2}$ can be calculated. The result ecreasing function of q. Equation plies that D is at a minimum for a n value of q which can be deter- by setting $dD/dq=0$.

ple 2

power spectra of x_α and δ are given

$$=\frac{C_1'}{(1-z^{-1})(1-z)} \tag{61}$$

$$=2\Delta \tag{62}$$

ation 61 represents independent ke variations in the operating situa- nd equation 62 can be explained as sult of independent measurements nean square error Δ for each meas- nt of period T_1. Determine D, q, $V(z)$.

on

result of the preceding section

$$\frac{(1-a)z^{-1}}{1-az^{-1}},\quad W(z)=\frac{1-a}{2b'q} \tag{63}$$

$$2+\frac{2b'^2q^2C_1'}{\Delta} \tag{64}$$

$$\overline{x_i)^2}=\frac{\Delta}{2b'^2q^2}\left(\frac{1}{a}-1\right) \tag{65}$$

ing q^2 in terms of a from equation

$$\frac{\Delta}{2C_1'}\frac{(1-a)^2}{a} \tag{66}$$

From equations 65 and 66, equation 67 can be written as

$$D=\frac{b'C_1'}{2(1-a)}+\frac{T_1\Delta}{2b'TC_1'}\frac{(1-a)^2}{a} \tag{67}$$

In equation 67, only a depends on q. For optimum q, dD/da is equated to zero, and the result is

$$\frac{a^2}{(1-a)^2(1-a^2)}=\frac{T_1\Delta}{Tb'^2C_1'^2}\equiv\eta \tag{68}$$

The second equality sign defines the parameter η.

For optimum choice of q, equation 67 can be written as

$$D=\frac{b'C_1'}{2}f_D \tag{69}$$

where

$$f_D=\frac{1+2a}{1-a^2} \tag{70}$$

Equation 66 can be rewritten as

$$q^2=\frac{TC_1'}{2T_1}\frac{a}{1-a^2}\equiv\frac{TC_1'}{2T_1}f_q \tag{71}$$

From equations 68, 70, and 71, η, f_D and f_q are calculated. The functions f_D and f_q are plotted against η in Fig. 7, and a is plotted against η in Fig. 8. In a design problem, η is calculated from given data. The minimum value of D together with the values of q and a to be used are read from these curves.

Analysis of Alternate Biasing Systems

For an alternate biasing system, let the actually applied x_i' be

$$x_i'=x_i+(-1)^iq \tag{72}$$

Then

$$m_s'=m_{opt}-\frac{b'}{2}[x_{\alpha i}-x_i-(-1)^iq]^2+\delta^i \tag{73}$$

The signal $(-1)^im_i'$ is used for adjustment of x_i, and will be denoted as u_i.

$$x_{i+1}=x_i+\sum_{j=0}^{j=i}w_{i-j}u_j \tag{74}$$

$$u_i=(-1)^im_i'=b'q(x_{\alpha i}-x_i)+(-1)^i\delta_i+(-1)^i\left[m_{opt}-\frac{b'}{2}q^2-\frac{b'}{2}(x_{\alpha i}-x_i)^2\right] \tag{75}$$

Except for the terms in the square bracket, equations 74 and 75 are identical with equations 6 and 7 and $b=b'q$. In the square bracket, m_{opt} is a slow-varying term, $b'/2\ q^2$ is a constant, $b'/2(x_{\alpha i}-x_i)^2$ is small and always positive. If these terms can be approximated by a constant C_2, the additional disturbance is

$$\delta_i'=(-1)^iC_2$$

and its z-transform is

$$\delta'(z)=\frac{C_2}{1+z^{-1}}$$

Suppose that a filtering factor $1+z^{-1}$ is introduced in $W(z)$. The disturbance δ' is filtered out except for the initial period $i=1$. Since the adaptive control loop must be stable, its impulse response is a vanishing function of t. The disturbance due to δ_i' reaches the vanishing point except for the first few sampling periods immediately after turning on the adaptive controller. Allowing a warm up period, one may assume that by introducing a filtering factor $1+z^{-1}$ in $W(z)$, the effect of the terms in the square bracket on D can be neglected. This point will be verified by a step-by-step calculation later in the paper.

Summarizing the foregoing, Fig. 2 gives a fairly good approximation of an alternate biasing system with the additional requirement that $1+z^{-1}$ is a factor in the numerator of $W(z)$.

To determine the optimum form of $W(z)$ let

$$K(z)=z^{-1}(1+z^{-1})F(z) \tag{76}$$

where $F(z)$ is an unknown function with

Table II. Comparison of Extremal-Seeking Systems

	Derivative Sensing Systems	Alternate Biasing
	$T=4T_1$	$T=10T_1$	$T=T_1$
C_1'	$0.2\ C_1$	$0.5\ C_1$	$0.05\ C_1$
η	0.625	0.04	40
f_D	2.20	1.36	7.05
f_q	0.51	0.17	2.15
a	0.42	0.17	0.79
q^2	$0.204\ C_1$	$0.42\ C_1$	$0.107\ C_1$
D	$0.22\ b'C_1$	$0.34\ b'C_1$	$0.182\ b'C_1$

Table III. Step-by-Step Calculation of an Alternate Biasing System

Case A: $x_i'=x+(-1)^i$

i	x_i	x_i'	m	$(-1)^i m$	$x_{i+1}-x_i$
0	0	1.0	−8	−8	−0.8
1	−0.8	−1.8	−32.7	32.7	1.83
2	1.83	2.83	−0.37	−0.37	4.5
3	4.53	3.53	0.779	−0.779	3.875
4	3.875	4.875	0.235	0.235	3.952
5	3.952	2.952	−0.106	0.106	3.971
6	3.971	4.971	0.058	0.058	3.985

Case B: $x_i'=x-(-1)^i$

x_i	x_i'	m	$(-1)^{i+1}m$	$x_{i+1}-x_i$
0	−1	−24	24	2.4
2.4	3.4	0.64	0.64	1.984
4.384	3.384	0.62	−0.62	−0.395
3.989	4.989	0.022	0.022	0.019
4.008	3.008	0.016	−0.016	−0.003

all its poles inside the unit circle. The point $z=-1$ is treated as a point outside the unit circle. The method of reference 12 gives:

$$K(z)=\frac{1+z^{-1}}{Y(z)}\left[\frac{\Phi_{x_\alpha x_\alpha}(z)}{(1+z^{-1})Y(z)^{-1}}\right]_+ \tag{77}$$

where $Y(z)$ is defined by

$$Y(z)Y(z^{-1})\equiv\Phi_{x_\alpha x_\alpha}(z)+\bar{\Phi}_{\delta\delta}(z)/b^{-2}q^2 \tag{78}$$

and $\bar{\Phi}_{\delta\delta}(z)$ is the sampled spectrum of $(-1)^i\delta_i$.

$$\overline{(x_{\alpha_i}-x_i)^2}=\frac{1}{2\pi j}\oint\left\{[1-K(z)]\times[1-K(z^{-1})]\Phi_{x_\alpha x_\alpha}(z)+\frac{K(z)K(z^{-1})}{b'^2q^2}\bar{\Phi}_{\delta\delta}(z)\right\}\frac{dz}{z} \tag{79}$$

$$D=\frac{b'}{2}\overline{(x_\alpha-x_i)^2}+\frac{b'}{2}q^2 \tag{80}$$

Example 3. Assuming the same power spectra as given in Example 2, determine optimum $W(z)$, q and D.

Solution. Since the successive samples of δ_i are uncorrelated

$$\bar{\Phi}_{\delta\delta}(z)=\Phi_{\delta\delta}(z)=\Delta$$

Equation 77 gives

$$K(z)=\frac{(1-a)z^{-1}(1+z^{-1})}{2(1-az^{-1})} \tag{81}$$

$$W(z)=\frac{(1-a)(1+z^{-1})}{2b'q\left(1+\frac{1-a}{2}z^{-1}\right)} \tag{82}$$

where

$$a+\frac{1}{a}=2+\frac{b'^2q^2C_1'}{\Delta} \tag{83}$$

Following the same steps as Example 2, the results are

$$\overline{(x_{\alpha_i}-x_i)^2}=\frac{5-a}{4(1-a)}C_1' \tag{84}$$

$$\eta\equiv\frac{\Delta}{C_1'^2b'^2}=\frac{a^2}{(1-a^2)(1-a)^2} \tag{85}$$

$$D=\frac{b'C_1'}{2}\left[\frac{5+8a-a^2}{4(1-a^2)}\right]=\frac{b'C_1'}{2}\left(f_D+\frac{1}{4}\right) \tag{86}$$

$$q^2=C_1'\cdot\frac{a}{1-a^2}=f_qC_1' \tag{87}$$

where f_D and f_q are the same functions as defined in equations 70 and 71. Given η, the values of f_D, f_q, and a can be read from Figs. 7 and 8.

Comparison of the Two Types of Extremal Seeking Systems

The two types of extremal seeking systems can be compared on the basis of equal $\Phi_{x_\alpha x_\alpha}(s)$, measuring interval T_i, and Δ. For a derivative sensing system, T is much larger than T_1 and will be assumed as at least $4T_1$ for the present purpose. For an alternate biasing system, $T=T_1$. For instance, let the following values be assumed:

$$\nu=0.025/T_1$$

$$\Delta=0.1b'^2C_1^2$$

The calculated results are listed in Table II, which shows clearly the advantage of an alternate biasing system. Not only is its deviation function D lower, the advantage of which is clear, but its lower value of test bias q is also a desirable feature.

A lower value of q means less disturbance to the system.

$$a=0.6 \qquad b'=2, \qquad q=1$$

$$K(z)=\frac{0.2z^{-1}(1+z-1)}{1-0.6\ z^{-1}} \tag{88}$$

$$W(z)=\frac{0.1(1+z^{-1})}{1+0.2z^{-1}} \tag{89}$$

$$m_i=1-x_i'^2, \quad i=-\infty\ldots-2,-1$$

$$m_i=1-(4-x_i')2, i=0,1,2\ldots \tag{90}$$

The change in x_α is expressed as

$$X_\alpha(z)=\frac{4}{1-z^{-1}} \tag{91}$$

There are two possibilities:

A. $x_i'=x_i+(-1)^i \qquad u_i=(-1)^i m_i$

B. $x_i'=x_i-(-1)^i \qquad u_i=(-1)^{i+1}m_i$

The adjustment equation is given $W(z)$ as

$$x_{i+1}-x_i=0.1(m_i+m_{i-1})-0.2(x_i-x_{i-1})$$

Table III is a step-by-step calculation equations 90, 92, and 93, which give illustration of how step-by-step adjustments are made in an alternate bias system. The total loss in m in the transient period is 40.1 for Case A and 22 Case B. The average value is 31.4. check of the linear approximation, integral value of loss in m is

$$\overline{(x_{\alpha_i}-x_i)^2}=\frac{1}{2\pi j}\times\oint\frac{16[1-K(z)][1-K(z^{-1})]}{(1-z^{-1})(1-z)}d.$$

The agreement is as good as ca expected.

Conclusions

A study is made on the optim utilization of measured data to r adjustments in a self-optimizing sys A basic assumption is that there value of the adjustable paramete which gives the desired result in e interval. The desired result is to ke certain performance parameter either prescribed value or at an unknown tremal value. While the value of not known to the designer, its power trum is assumed to be known. point is deemed essential. If one not know how fast the operating situa varies, there is no possible optimiza of the adjustment procedure. Assu constant situation, the more mea ments one averages, the less error gets. Obviously this is not the ans

Given the power spectrum of probable error in each measurement, the dependence of the perform variable on the adjustable paramet the minimum value of D and a s weighting factors w_i can be obta The increment in x is computed weighted sum of the present and pre measurements. The set of weig factors w_1 gives optimum adjustmen a nominal situation, and near opti adjustment in general.

Two types of extremal seeking sys are studied: the derivative sensing tems and the alternate biasing syst In addition to determining the mini D and the weighting factors, a me for determining the optimum of test bias is also developed.

ace-wise, the alternate biasing sys-
re found to be more satisfactory
he derivative sensing systems.

ences

NCIPLES OF OPTIMALIZING CONTROL SYS-
ND AN APPLICATION TO THE INTERNAL
TION ENGINE, C. S. Draper, Y. T. Li.
n Society of Mechanical Engineers, New
. Y., Sept. 1951.

STOCHASTIC APPROXIMATION METHOD, H.
, S. Nunro. *Annals of Mathematical*
s, Michigan State College, East Lansing,
vol. 22, Sept. 1951, p. 400.

CHASTIC ESTIMATION OF THE MAXIMUM OF
ESSION FUNCTION, J. Keifer, J. Wolfowitz.
ol. 23, Sept. 1952, p. 462.

F-OPTIMIZING SERVO SYSTEMS WITH RAN-
DOM INPUTS, E. G. C. Burt. *Report*, Seminar on Non-Linear Control Problems, Cambridge University, Cambridge, England, Sept. 1954.

5. DISTURBANCE-RESPONSE FEEDBACK—A NEW CONTROL CONCEPT, J. B. Reswick. *Transactions*, American Society of Mechanical Engineers, vol. 78, Jan. 1956, p. 153.

6. OPTIMIZING CONTROL SYSTEMS, R. L. Cosgriff, R. A. Emerling. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, Mar. 1958, pp. 13–16.

7. DESIGN OF A SELF-OPTIMIZING CONTROL SYSTEM, R. E. Kalman. *Paper no. 57-IRD-2*, American Society of Mechanical Engineers, Apr. 1957.

8. ADAPTIVE SERVOMECHANISMS, R. F. Drenick, R. A. Shahbender. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, Nov. 1957, pp. 286–92.

9. OPTIMALIZING CONTROL-DESIGN OF A FULLY AUTOMATIC CRUISE CONTROL SYSTEM FOR TURBO-JET AIRCRAFT, W. K. Genthe. *Wescon Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957.

10. CONTINUOUS MEASUREMENT OF CHARACTERISTICS OF SYSTEMS WITH RANDOM INPUTS: A STEP TOWARD SELF OPTIMIZING CONTROL, T. P. Goodman, R. H. Hillsley. *Paper no. 58-IRD-5*, American Society of Mechanical Engineers, 1958.

11. A SELF-ADJUSTING SYSTEM FOR OPTIMUM DYNAMIC PERFORMANCE, G. W. Anderson, J. A. Aseltine, A. R. Mancini, C. W. Sarture. *National Convention Record*, Institute of Radio Engineers, pt. 4, 1958.

12. STATISTICAL DESIGN THEORY FOR STRICTLY DIGITAL SAMPLED-DATA SYSTEMS, S. S. L. Chang. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 702–09.

13. SAMPLED-DATA PROCESSING TECHNIQUES FOR FEEDBACK CONTROL SYSTEMS, A. R. Bergen, J. R. Ragazzini. *Ibid.*, pt. II (*Applications and Industry*), vol. 73, Nov. 1954, pp. 236–47.

ERRATA

The following errata in previous volumes have just been brought to our attention.

"Comparison of Backlash and Hysteresis Effects in Second-Order Feedback Systems" by L. M. Vallese, published in AIEE Transactions, part II (Applications and Industry), volume 75, September 1956, pages 240–43.

The author wishes to make the following corrections (see reference 1). Equation 26 on page 242 should read:

$$\frac{dA_x}{dt}=\frac{-1}{2\pi\omega}\left(A_x a_1\Omega\pi-4x_0ka_2+\frac{4x_0^2ka_2}{A_x}\right) \quad (26)$$

Equation 32(A) on page 243 should read:

$$\xi\lesssim\frac{\pi\Omega_1}{4\alpha a_1(1-\alpha)} \quad (32A)$$

REFERENCE

1. BACKLASH AND HYSTERESIS EFFECTS IN AUTOMATION SYSTEMS, L. M. Vallese. *Journal*, British Institution of Radio Engineers, London, England, vol. 17, June 1957, p. 315.

"Selection of Buffer Reactors and Synchronous Condensers on Power Systems Supplying Arc-Furnace Loads" by C. Concordia, L. G. Levoy, and C. H. Thomas, published in AIEE Transactions, part II (Applications and Industry), volume 76, July 1957, pages 123–35.

The authors point out that the illustrations published as Fig. 16 on page 131 and Fig. 17 on page 132 should be interchanged.

Power Apparatus and Systems—June 1960

60-98 Calculation of Stray Load Losses in D-C Machinery.....Erdelyi . . . 129
60-53 Conductors for EHV Transmission.....Abetti, Lindh, Simmons . . . 138
60-95 RI Studies for EHV Lines in Ontario.....Reichman, Leslie . . . 153
60-83 Arcing Fault Protection for Low-Voltage Systems.....Kaufmann, Page . . . 160
60-90 Electric Stress Grading of Insulator Strings.....Denholm . . . 167
60-3 Temporary Paralysis Following "Freezing" to a Wire.....Dalziel . . . 174
60-94 Control of the Thermal Environment of Buried Cables.....Schmill . . . 175
60-125 Impulse Testing of Rotating A-C Machines.....Committee Report . . . 182
60-32 Survey of Induction Motor Protection.....Committee Report . . . 188
60-141 Characteristics of Power Cable Shielding.....Mitropoulos, Fogel, Tang . . . 192

Distribution System Planning Through Optimized Design.....
.....Lawrence, Reps, Patton
60-177 I—Distribution Transformers and Secondaries..... . . . 199
60-178 II—Comparative Economics of System Voltages..... . . . 204

60-138 Using the Digital Computer in Power System Design.....Cook, Patton . . . 223
60-104 Computer for Electric Load Monitoring.....Thomas, Gustafson, Foster . . . 235
60-92 Switching Surges on Transformer-Terminated Line...Johnson, Schultz . . . 241
60-155 Pilot-Wire Protection of Transmission Lines.....Neher . . . 245
60-154 Effect of Rectifier Power Supply on Large D-C Motors.....Dunaiski . . . 253
60-41 End Component of Zero-Sequence Reactance of A-C Machines....Smith . . . 259
60-164 Coupling-Capacitor Potential Device.....Norimatsu, Uyeda . . . 264
60-93 Comparison of European and American Wet Flashover Test...Kingsbury . . . 271
60-169 Commutator Primitive for Generalized Machine Theory...Barton, Jones . . . 277
60-185 International Standardization of EHV Transmission....Sporn, St. Clair . . . 281
60-82 New Method of Making Transmission-Loss Formulas.....George . . . 287
60-191 Measuring the Lightning Strength of High-Voltage Insulators.....Lantz . . . 298
60-192 R-F on AEP 345-Kv Line.....Barthold, LaForest, Schlomann, Trebby . . . 303
60-186 Project EHV—Preliminary Corona Investigations.....Foley, Olsen . . . 310
60-175 Audible Noise from Power Transformers.....Schulz, Ringlee . . . 316
60-180 Digital Calculation of Line-to-Ground Short Circuits.....El-Abiad . . . 323
60-205 Insulation Co-ordination of Circuit Breakers.....Streater, Wilson . . . 333
60-239 Broken Wire Assumptions.....Committee Report . . . 344

Conference Papers Open for Discussion

The conference papers listed below have been accepted for AIEE Transactions and are now open for written discussion until August 29. Duplicate double-spaced typewritten copies for each discussion should be sent to Edward C. Day, Assistant Secretary for Technical Papers, American Institute of Electrical Engineers, 33 West 39th Street, New York 18, N.Y., on or before August 29.

Preprints may be purchased at 50¢ each to members; $1.00 each to nonmembers if accompanied by remittance or coupons. Please order by number and send remittance to:

AIEE Order Department
33 West 39th Street
New York 18, N. Y.

58-1362 A Digital Nuclear Reactor Control System....Gyftopoulos, Coble
60-23 Economic Trends Make Railroad Electrification Inevitable.Curtis
60-601 Air-Cleaning Features for Traction Equipment.....Lessmann

AIEE PUBLICATIONS

	Member Prices	Nonmember Prices: Basic Prices*†	Nonmember Prices: Extra Postage for Foreign Subscription
Electrical Engineering			
Official monthly publication containing articles of broad interest, technical papers, digests, and news sections: Institute Activities, Current Interest, New Products, Industrial Notes, and Trade Literature. Automatically sent to all members and enrolled students in consideration of payment of dues. (Members may not reduce the amount of their dues payment by reason of nonsubscription.) Additional subscriptions are available at the nonmember rates.		annually $12*	$1.00
		Single copies $1.50*	
Bimonthly Publications			
Containing all officially approved technical papers collated with discussion (if any) in three broad fields of subject matter as follows:			
	annually	annually	
Communication and Electronics	$5.00	$8.00*	$0.75
Applications and Industry	$5.00	$8.00*	$0.75
Power Apparatus and Systems	$5.00	$8.00*	$0.75
Each member may subscribe to any one, two, or all three bimonthly publications at the rate of $5.00 each per year. A second subscription to any or all of the bimonthly publications may be obtained at the nonmember rate of $8.00 each per year.			
Single copies may be obtained when available.	$1.50 each	$1.50* each	
AIEE Transactions			
An annual volume in three parts containing all officially approved technical papers with discussions corresponding to six issues of the bimonthly publication of the same name bound in cloth with a stiff cover.			
	annually	annually	
Part I Communication and Electronics	$4.00	$8.00*	$0.75
Part II Applications and Industry	$4.00	$8.00*	$0.75
Part III Power Apparatus and Systems	$4.00	$8.00*	$0.75
Annual subscription to all three parts (beginning with vol. 77 for 1958).	$10.00	$20.00*	$2.25
Annual subscription to any two parts.		$15.00*	$1.50
AIEE Standards			
Listing of Standards, test codes, and reports with prices furnished on request.			
Special Publications			
Committee reports on special subjects, bibliographies, surveys, and papers and discussions of some specialized technical conferences, as announced in ELECTRICAL ENGINEERING.			

*Discount 25% of basic nonmember prices to college and public libraries. Publishers and subscription agencies 15% of basic nonmember prices. For available discounts on Standards and special publications, obtain price lists from Order Department at Headquarters.

†Foreign prices payable New York exchange

Send all orders to:

Order Department
American Institute of Electrical Engineers
33 West 39th Street, New York 18, N. Y.